

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

Run on: July 8, 2002, 08:19:12 ; Search time 51.87 Seconds

42.828 Million cell updates/sec

Title: US-09-582-296-2

Sequence: 1 APAVVMGDAESFGAIAHGGL 20

Scoring table: BLOSUM62

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

### Listing first 45 summaries

Database : A\_Geneseq\_032802:\*

1:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1980.DAT.*
2:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1981.DAT.*
3:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1982.DAT.*
4:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1983.DAT.*
5:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1984.DAT.*
6:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1985.DAT.*
7:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1986.DAT.*
8:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1987.DAT.*
9:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1988.DAT.*
10:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1989.DAT.*
11:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1990.DAT.*
12:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1991.DAT.*
13:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1992.DAT.*
14:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1993.DAT.*
15:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1994.DAT.*
16:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1995.DAT.*
17:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1996.DAT.*
18:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1997.DAT.*
19:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1998.DAT.*
20:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA1999.DAT.*
21:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA2000.DAT.*
22:	/SIDS1/gcgdata/hold-genseq/genseqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100	0	20	AAI27011	N-terminal fragment
2	69	68	3	18	AAI27010	N-terminal fragment
3	51.5	51	0	22	AAI27010	C glutamicum prote
4	48	47	5	22	AAU01898	Mycobacterium tube
5	45.5	45	0	899	ABB65488	Drosophila melanog
6	45.5	45	0	899	ABB65489	Drosophila melanog
7	45	44	6	617	AAI49925	Consensus haemaggl
8	44.5	44	1	474	AAU3631	Pseudomonas aerugi
9	43	42	6	130	AAI29756	Rice glutaredoxin
10	43	42	6	317	AAI74126	Human prostate tum
11	43	42	6	330	AAW94066	Human DNase-like pr

	12	43	42.6	524	20	AAR94753	Mutant H protein
	13	43	42.6	617	14	AAR34541	H protein of atlen
	14	43	42.6	617	14	AAR42387	Moraten haemagglut
	15	43	42.6	617	14	AAR42388	Consensus haemagg
	16	43	42.6	617	14	AAR42389	San Diego haemagg
	17	43	42.6	617	14	AAR42390	Chicago 1 haemaggi
	18	43	42.6	617	14	AAR42391	Chicago 2 haemaggi
	19	43	42.6	617	14	AAR42392	McI haemagglutinln
	20	43	42.6	617	14	AAR42393	JM haemagglutinin
	21	43	42.6	617	20	AAR94752	Mutant measles vir
	22	42.5	42.1	385	22	AAG90134	C glutamincum prote
	23	42	41.6	106	22	AAG01349	Novel human diagn
	24	42	41.6	188	22	AAB94158	Human protein sequ
	25	42	41.6	180	22	AAG66383	Human partial olfa
	26	42	41.6	216	22	AAG72004	Human olfactory re
	27	42	41.6	216	22	AAG72923	Human olfactory re
	28	42	41.6	216	22	AAG72924	Human olfactory re
	29	42	41.6	216	22	AAG72926	Human olfactory re
	30	42	41.6	228	22	ABG01309	Novel human diagno
	31	42	41.6	239	22	ABG01350	Novel human diagno
	32	42	41.6	301	22	AAG72927	Human olfactory re
	33	42	41.6	305	22	AAG66387	Human partial olfa
	34	42	41.6	305	22	AAG66389	Human partial olfa
	35	42	41.6	317	22	AAG71408	Human olfactory re
	36	42	41.6	317	22	AAG72330	Human OR-like poly
	37	42	41.6	317	22	AAG72925	Human olfactory re
	38	42	41.6	317	22	AAG72977	Olfactory receptor
	39	42	41.6	584	22	ABH68361	Drosophila melanog
	40	42	41.6	608	22	AAG91551	C glutamincum prote
	41	41	40.6	116	21	AAY50927	Human fetal brain
	42	41	40.6	131	21	AAB57124	Human prostate can
	43	41	40.6	143	21	AAB40458	Human ORFX ORF222
	44	41	40.6	185	21	AAG52283	Arabidopsis thaliaa
	45	41	40.6	185	22	AAG91571	C glutamincum prote

ALIGNMENTS

RESULT	1	
ID	AAY27011	
XX	AAY27011 standard; peptide; 20 AA.	
XX	AAY27011;	
DT	24-SEP-1999	(first entry)
XX	N-terminal fragment of anti-freeze protein from Lichen.	
DE		
XX	Antifreeze protein; AFP; Lichen; frozen food product; pasteurisation;	
KW	Sterilisation; freezing; frozen confectionery product; N-terminal.	
XX		
OS	Umbilicaria antarctica.	
XX		
NN	WO9937673-A2.	-
XX		
PD	29-JUL-1999.	
XX		
PF	23-DEC-1998;	98WO-EF08554.
PR	22-JAN-1998;	98GB-0001420.
XX		
PA	(UNIL ) UNILEVER NV.	
EA	(UNIL ) UNILEVER PLC.	
P1	Byass LJ, Sidebottom CM,	Smallwood MF;
DR	WPI; 1999-444595/37.	
PT	New isolated antifreeze protein obtained from Lichen, used for the	
CT	preparation of food products, particularly frozen confectionery	
XX	products	



DT	26-MAR-2002	(first entry)
XX		
XX		
DE	Drosophila melanogaster	polypeptide seq ID NO 23256.
XX		
KW	Drosophila; developmental biology; cell signalling; insecticide;	
KW	pharmaceutical.	
XX		
XX	Drosophila melanogaster.	

EN NO2001/1042 NZ.  
 XX  
 PD 27-SEP-2001.  
 XX

PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX  
PI Venter JC, Adams M, Li PM  
XX

DR	N-PSDB; ABL09591.
XX	
PT	New isolated nucleic acid de
PT	genes from Drosophila and fc
PT	interactions -

PS	Disclosure; SEQ ID NO 23256;
XX	
CC	The invention relates to an
CC	capable of detecting 1000 or

CC discloses genomic DNA sequences (ABL01840-ABL161755 (ABB57737-ABB72072)).

CC at ftp.wipo.int/pub/publishe  
XX  
SQ Sequence 899 AA;

	Query Match	45.0%;
	Best Local Similarity	55.0%;
	Matches 11; Conservative	
QY	2 PAVYMGDA-ESFAIANGGL 20	
	:   :	
Db	82 pwiavgkltqisfglaingql 10	

## RESULT 6

XX		ABBS65489;
XX		
AC		
XX	26-MAR-2002	(first entry)
DT		
XX		Drosophila melanogaster poly
DE		
XX		

XX	Pharmaceutical.
XX	
OS	<i>Drosophila melanogaster</i> .
XX	
PN	W0200171042-A2.
XX	
AD	27-SEP-2001.
	27-AD





Query Match 42.6%; Score 43; DB 22; Length 130;  
 Best Local Similarity 55.0%; Pred. No. 11;  
 Matches 11; Conservative 2; Mismatches 5; Indels 2; Gaps 1;  
 OY 2 PAV--VMGDAESFGAIAHGC 19  
 111 1111 1111  
 Db 70 pavhevagaealagvva99 89

RESULT 10  
 AAY74126  
 ID AAY74126 standard; Protein; 317 AA.  
 XX  
 AC AAY74126;  
 XX  
 DT 14-MAR-2000 (first entry)  
 XX  
 DE Human prostate tumor EST fragment derived protein #313.  
 XX  
 KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;  
 XX treatment.  
 XX Homo sapiens.  
 XX OS  
 XX PN DE19820190-A1.  
 XX PD 04-NOV-1999.  
 XX PF 28-APR-1998; 98DE-1020190.  
 XX PR 28-APR-1998; 98DE-1020190.  
 XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
 XX PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
 XX DR WPI: 1999-621386/54.  
 XX DR N-PSDB; AA52961.  
 XX PT New human nucleic acid sequences from pancreatic tumors, and related  
 XX proteins -  
 XX PS Claim 23; Page 439; 502pp; German.  
 XX CC This invention describes novel polypeptides and their encoding nucleic  
 XX acids derived from human pancreatic tumor tissue which have cytostatic  
 XX activity. The sequences are also useful in producing pharmaceutical  
 XX compositions for treatment of pancreatic tumors. AAY73814-Y74252  
 XX CC represent protein fragments encoded by the human pancreatic tumor cDNA  
 XX library derived expressed sequence tag (EST) sequences represented in  
 XX CC AA52858-253014.  
 XX SQ Sequence 317 AA;  
 Query Match 42.6%; Score 43; DB 20; Length 317;  
 Best Local Similarity 70.0%; Pred. No. 32;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 OY 11 SFGAIAHAGGL 20  
 1111111111  
 Db 164 sfsglghgyl 173  
 RESULT 11  
 AAM94066  
 ID AAM94066 standard; Protein; 330 AA.  
 XX  
 AC AAM94066;  
 XX  
 DT 09-APR-1999 (first entry)  
 XX  
 DE Human DnaJ-like protein, HSPJ2.

XX  
 KW DnaJ-like protein; heat shock protein; HSPJ1; HSPJ2; cancer; leukaemia;  
 KW immune disorder; inflammation; tissue damage; diabetes; wound healing;  
 KW chromosome mapping.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key location/Qualifiers  
 FT Misc-difference 239  
 FT /label= unknown  
 FT /note= "encoded by GST"  
 FT Misc-difference 272  
 FT /label= unknown  
 FT /note= "encoded by GNG"  
 FT Misc-difference 290  
 FT /label= unknown  
 FT /note= "encoded by GNG"  
 FT Misc-difference 291  
 FT /label= unknown  
 FT /note= "encoded by NCC"  
 FT Misc-difference 293  
 FT /label= unknown  
 FT /note= "encoded by NNC"  
 FT Misc-difference 298  
 FT /label= unknown  
 FT /note= "encoded by NCA"  
 FT Misc-difference 301  
 FT /label= unknown  
 FT /note= "encoded by TTN"  
 FT Misc-difference 302  
 FT /label= unknown  
 FT /note= "encoded by NAA"  
 FT Misc-difference 305  
 FT /label= unknown  
 FT /note= "encoded by NGC"  
 FT Misc-difference 309  
 FT /label= unknown  
 FT /note= "encoded by NAA"  
 PN W09855509-A2.  
 XX  
 PD 10-DEC-1998.  
 XX  
 PF 02-JUN-1998; 98WO-US11182.  
 XX  
 PR 03-JUN-1997; 97US-0868288.  
 XX  
 PA (INCY-) INCYTE PHARM INC.  
 XX  
 PI Au-Young J, Bandman O, Lal P;  
 XX  
 DR WPI: 1999-070259/06.  
 DR N-PSDB; AAX06100.  
 XX  
 PT New nucleic acid encoding human DnaJ-like proteins - for diagnosis,  
 PT treatment and prevention of cancer, immune disorders and  
 PT inflammation  
 XX  
 PS Claim 22; Fig 3A-D; 73pp; English.  
 XX  
 CC This represents a human DnaJ-like protein, HSPJ2. The invention provides  
 CC two human DnaJ-like proteins which are heat shock proteins J1 and J2  
 CC (HSPJ1 and HSPJ2) and nucleic acid sequences encoding the proteins. Host  
 CC cells containing a vector comprising the nucleic acids are used for the  
 CC production of recombinant proteins. Recombinant HSPJ1 and HSPJ2 are used  
 CC to raise Ab, therapeutically and to screen for specific binding agents.  
 CC Antagonists are used to treat or prevent a wide variety of solid cancers,  
 CC leukaemia and lymphoma; immune disorders (typical of many disclosed are  
 CC acquired immune deficiency syndrome, allergy, asthma, Crohn's disease,  
 CC diabetes mellitus, multiple sclerosis, rheumatoid arthritis, microbial  
 CC and parasitic infections) and inflammation. Agonists may be used to treat  
 CC or prevent tissue damage, e.g. in cases of ankylosing spondylitis, heart  
 CC attack, ischaemia, UV exposure, insulin-dependent diabetes and for wound

DE Moraten haemagglutinin.

XX Hemagglutinin; HA; fusion glycoprotein; wild-type;  
 KM measles virus; vaccine; infection; consensus polypeptide.  
 XX  
 OS Measles virus strain Moraten.  
 PN WO9321325-A.  
 XX  
 PD 28-OCT-1993.  
 PF  
 PF 08-APR-1993; 93WO-US03209.  
 XX  
 PR 08-APR-1992; 92US-0866033.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 XX  
 PI Bellini WJ, Rota JS;  
 XX  
 DR WPI; 1993-351735/44.  
 DR N-PSDB; AA051088.  
 XX  
 PT Hemagglutinin and fusion glycoprotein of several wild-type  
 PT measles strains - used to construct vaccines for measles  
 PT infection  
 XX  
 PS  
 PS Disclosure; Page 20-22; 119pp; English.  
 XX  
 CC HA (AA051088-94) and fusion glycoprotein (AA051095-97) sequences of  
 CC several wild-type measles strains are given. Shared amino acid  
 CC variations in wild-type measles glycoproteins are identified in five  
 CC wild-type measles viruses. A consensus polypeptide, the amino acid  
 CC sequence of which reflects variation common to more than one wild-  
 CC type strain, is the basis for constructing live attenuated vaccines  
 CC or recombinant vaccines to replace older, less efficacious vaccines  
 CC Immunological reagents useful in differentiating wild-type measles  
 CC strains from other known strains can also be produced.  
 XX  
 XX Sequence 617 AA: .

Query Match	42.6%	Score 43;	DB 14;	Length 617;
Best Local Similarity	50.0%;	Pred. No. 69;		
Matches	8;	Conservative	3;	Mismatches 5;
				Indels 0;
				Gaps 0;
QY	5	VMGDAESFGAIAHGCL	20	
	1:	1:11	1:1	1:
DB	584	VLADESGNHLHSGM	599	

RESULT	15
AA#42388	
ID	AA#42388 standard; Protein; 617 AA

AC AAR42388-

DT 13-MAY-1994 (first entry)

DE Consensus haemagglutinin sequence.

KW Haemagglutinin; HA; fusion glycoprotein; wild-type;  
measles virus; vaccine; infection; consensus polypeptide

OS Measles virus.

PN W09321325-A.

PD 28-OCT-1993

PF 08-APR-1993; 93WO-US03209.

PR 08-APR-1992; 92US-0866033.

PA (USSH) US DEPT HEALTH &amp; HUMAN SERVICE.

XX Bellini WJ, Rota JS,  
PI  
XX WPI; 1993-351735/44.  
DR N-PSDB; AAQ51089.  
VY

PT Haemagglutinin and fusion glycoprotein of several wild-type measles strains - used to construct vaccines for measles

PT infection

PS Disclosure; Page 25-28; 119pp; English

CC HA (AA51088-94) and fusion glycoprotein (AA051095-97) sequences of  
CC several wild-type measles strains are given. Shared amino acid  
CC variations in wild-type measles glycoproteins are identified in five  
CC wild-type measles viruses. A consensus polypeptide, the amino acid  
CC sequence of which reflects variation common to more than one wild-  
CC type strain, is the basis for constructing live attenuated vaccines  
CC or recombinant vaccines to replace older, less efficacious vaccines  
CC Immunological readouts useful in differentiating wild-type measles  
CC strains from other known strains can also be produced.

Sequence	617 AA;
5Q	

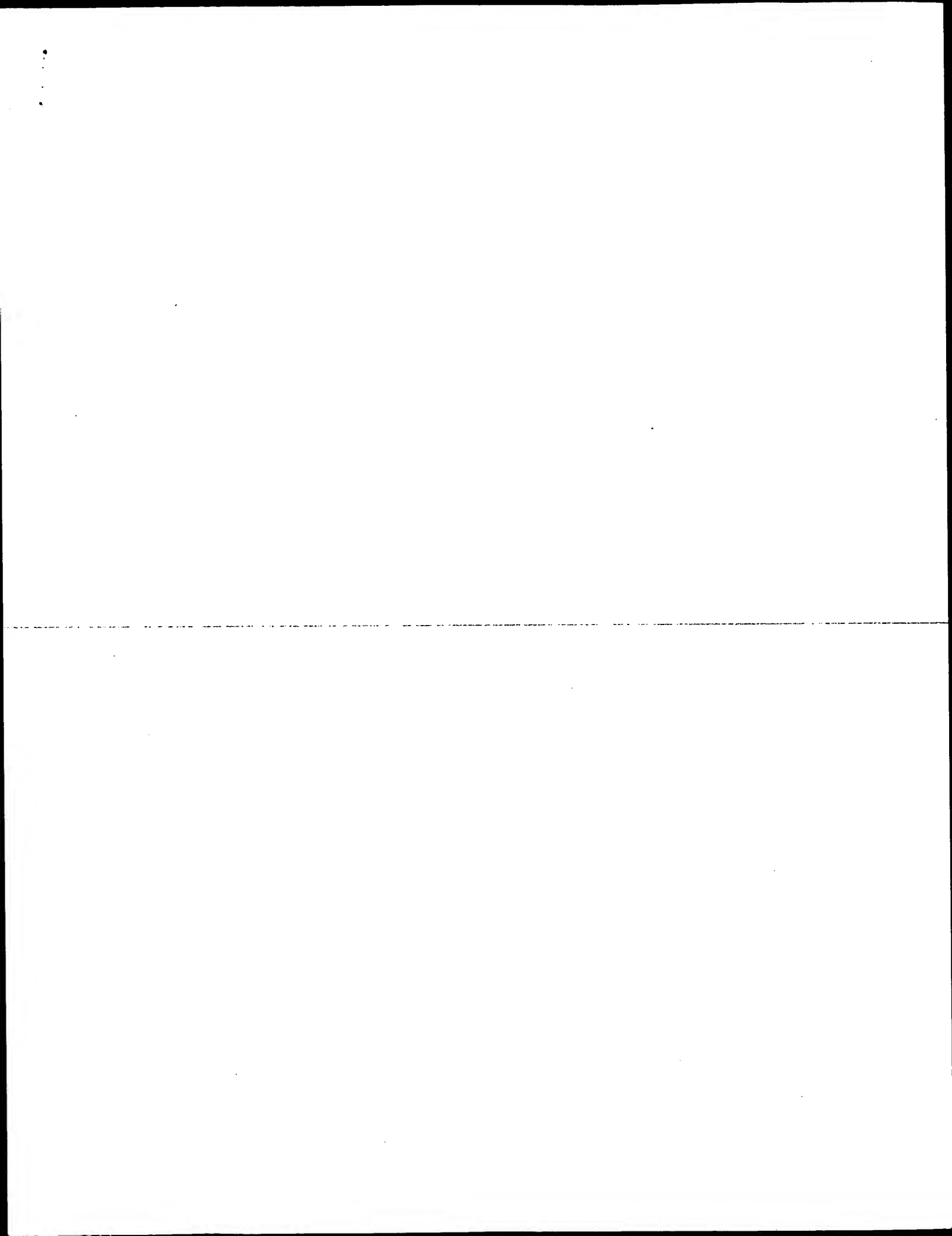
Query Match	42.6%;	Score 43;	DB 14;	Length 617;
Best Local Similarity	50.0%;	Pred. No. 69;		
Matches	8;	Conservative	3;	Mismatches 5;
				Indels 0;
				Gaps 0.

```
QY      5 VMGDAESFGAIAHGGL 20      |:|:| | | |:
Db      584 vladesesgghithsgm 599
```

Search completed: July 8, 2002, 08:19:13  
Job time: 578 sec

Mon Jul 8 09:33:16 2002

us-09-582-296-2.rag



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:17:25 ; Search time 41.68 Seconds  
(without alignments)  
74.710 Million cell updates/sec

Title: US-09-582-296-1  
Perfect score: 98  
Sequence: 1 APAMDAESFGAIGHGL 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 segs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_protist:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	57.1	145	2	09K4B2
2	50	51.0	963	13	042110
3	49	50.0	461	11	09DC17
4	49	50.0	461	11	09D728
5	49	50.0	461	11	099KR8
6	49	50.0	467	4	09BTV2
7	49	50.0	467	4	09UJM5
8	48	49.0	116	11	09D242
9	48	49.0	452	11	09DD22
10	48	49.0	572	11	099LJ1
11	48	49.0	572	11	042728
12	47	48.0	583	16	09PHM7
13	46	46.9	341	4	09P133
14	46	46.9	400	4	09NT44
15	46	46.9	628	4	09HB63
16	46	46.9	628	4	09B2P1

Seq ID no 1

17	45	45.9	363	16	091505	091505 pseudomonas
18	45	45.9	384	16	069579	069579 mycobacteri
19	45	45.9	449	16	09WE92	09WE92 thermotoga
20	44	44.9	281	16	098L72	098L72 rhizobium 1
21	44	44.9	421	16	092502	092502 rhizobium m
22	44	44.9	517	16	09C3A0	09C3A0 lactococcus
23	43.5	44.4	258	16	007601	007601 bacillus su
24	43.5	44.4	791	2	09L115	09L115 streptomyce
25	43	43.9	212	10	09ARX5	09ARX5 oryza sativ
26	43	43.9	246	16	0987Y7	0987Y7 rhizobium 1
27	43	43.9	250	4	09SH11	09SH11 homo sapien
28	43	43.9	250	4	096159	096159 homo sapien
29	43	43.9	263	10	09FR70	09FR70 oryza sativ
30	43	43.9	425	16	098L43	098L43 rhizobium 1
31	42.5	43.4	769	2	09R009	09R009 caulobacter
32	42.5	43.4	769	16	09A2S2	09A2S2 caulobacter
33	42.5	43.4	1222	5	09VF52	09VF52 dirosophila
34	42	42.9	334	17	029369	029369 archaeoglob
35	42	42.9	483	16	09RTM0	09RTM0 deinococcus
36	42	42.9	492	2	051065	051065 neisseria g
37	42	42.9	541	3	09P6T4	09P6T4 neurospora
38	42	42.9	547	10	09MA11	09MA11 arabidopsis
39	42	42.9	617	13	090ZF6	090ZF6 oryzae lat
40	42	42.9	700	13	091372	091372 xenopus lae
41	42	42.9	899	5	09VNO0	09VNO0 dirosophila
42	42	42.9	1486	16	09CG24	09CG24 lactococcus
43	41.5	42.3	233	16	098A01	098A01 rhizobium 1
44	41.5	42.3	474	12	089542	089542 bovine herp
45	41.5	42.3	518	2	09RKZ7	09RKZ7 streptomyce

## ALIGNMENTS

RESULT 1	ID	Q9K4B2	PRELIMINARY;	PRT;	145 AA.
Q9K4B2	AC	Q9K4B2;			
DT	01-OCT-2000	(TREMBLrel. 15, Created)			
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)			
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)			
DE	HYPOTHETICAL	15.4 KDA PROTEIN.			
GN	SC7E4.04C.				
OS	Streptomyces coelicolor.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
OX	NCBI_TaxID=1902;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-A3(2);				
RA	Seeger K.J., Harris D.;				
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-A3(2);				
RA	Cerdeno A.M., Parkhill J., Barrell B.G., Rajadream M.A.;				
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-A3(2);				
RX	MEDLINE=97000351; PubMed=8843436;				
RA	Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.;				
RL	Kinashi H., Hopwood D.A.;				
RT	"A set of ordered cosmids and a detailed genetic and physical map for				
RT	the 8 Mb Streptomyces coelicolor A3(2) chromosome.";				
RL	Mol. Microbiol. 21:77-96(1996).				
DR	EMBL; AL359214; CAB94592.1; -				
DR	InterPro: IPR000644; CBS.				
DR	Pfam: PF00571; CBS; 2.				
DR	SMART; SM00116; CBS; 2.				
KW	Hypothetical protein.				
SO	SEQUENCE 145 AA; 15353 MW; FAE58BBA18E358E5 CRC64;				

```

Query Match          57.1%; Score 56; DB 2; Length 145;
Best Local Similarity 58.8%; Pred. NO. 0.2;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 APAMDAESFGAIAHG 17
    |||  |  |||  |||
DB 90 APTWVEAGAGAMAHG 106

RESULT 2
Q02110 PRELIMINARY; PRT; 963 AA.
AC 042110;
AC 042113;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ORF2.
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Testudines; Pleurodira; Chelidae; Platemys.
OX NCBI_TaxID=45338;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSDON-RETROTRANSDON CRI-LIKE LINE (PSCR1);
RX MEDLINE=98066352; PubMed=9402732;
RA Kajikawa M., Ohsima K., Okada N.;
RT "Determination of the turtle CRI element encodes a protein with a novel
RT zinc finger motif.";
RL MOL. BIOL. EVOL. 14:1206-1217(1997).
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00078; rvt, 1.
KM RNA-directed DNA polymerase.
SQ SEQUENCE 963 AA; 109601 MW; 2842A81F9B6487D CRC64;

Query Match          51.0%; Score 50; DB 13; Length 963;
Best Local Similarity 50.0%; Pred. NO. 17;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAMMDAESFGAIAH 15
    |||  |  |||  |
DB 333 PAMLNKXKFGELKH 346

RESULT 3
Q09DCL7 PRELIMINARY; PRT; 461 AA.
AC 09DCL7;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 0610025011R1K.
GN 0610025011R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

```

```

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK002675; BAB22277.1;
DR MGD; MGI:1914098; 0610025011R1K.
DR InterPro: IPR000933; Alpha.L.fucos.
DR Pfam: PF01120; Alpha.L.fucos. 1.
DR PRINTS; PR00741; GUTHRLASE29.
DR PROSITE; PS00385; ALPHA_L_FUCOSIDASE; UNKNOWN_1.
SQ SEQUENCE 461 AA; 53631 MW; 1E233114026A3C1F CRC64;

Query Match          50.0%; Score 49; DB 11; Length 461;
Best Local Similarity 47.1%; Pred. NO. 10;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 PAMMDAESFGAIAHGL 18
    |||  |  |||  |
DB 41 PAMFDAQKFGFIHMGV 57

RESULT 4
Q09D728 PRELIMINARY; PRT; 461 AA.
AC 09D728;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 0610025011R1K.
GN 0610025011R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK008653; BAB25809.1;
DR MGD; MGI:1914098; 0610025011R1K.
DR InterPro: IPR000933; Alpha.L.fucos.
DR Pfam: PF01120; Alpha.L.fucos. 1.
DR PRINTS; PR00741; GUTHRLASE29.
DR PROSITE; PS00385; ALPHA_L_FUCOSIDASE; UNKNOWN_1.
SQ SEQUENCE 461 AA; 53530 MW; 25754D6F0F24D09B CRC64;

```

RESULT	6		
09BTV2			
ID	09BTV2	PRELIMINARY;	PR1; 467 AA.
AC	09BTV2;		
DT	01-JUN-2001 (TEMBLrel. 17, Created)		
DT	01-JUN-2001 (TEMBLrel. 17, last sequence update)		
DT	01-DEC-2001 (TEMBLrel. 19, last annotation update)		
DE	SIMILAR TO FUCOSIDASE, ALPHA-L-1, TISSUE.		
OS	Homo sapiens (Human)		
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=PLACENTA, CHORIOCARCINOMA;		
RA	Strausberg R.;		
RL	Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.		
DR	EMBL: BC003060; AAH03060.1; -		
DR	InterPro: IPR000933; Alpha_L_fucos.		
DR	Pfam: PF01120; Alpha_L_fucos. 1.		
DR	PRINTS: PR00741; G1HFDRLASE29.		
DR	PROSITE: PS00385; ALPHA_L_FUCOSIDASE; UNKNOWN.1.		
SO	SEQUENCE 467 AA; 54060 MW; F33BD0A9081BD85 CRC64;		

RESULT	8			
09D242				
ID	09D242	PRELIMINARY:	PRT:	116 AA.
AC	09D242:			
DT	01-JUN-2001	(TREMBLrel, 17, Created)		
DT	01-JUN-2001	(TREMBLrel, 17, Last sequence update)		
DT	01-JUN-2001	(TREMBLrel, 17, Last annotation update)		
DE	9530055J05RIK	PROTEIN.		
OS	9530055J05RIK			
GN	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10990;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=URINARY BLADDER.			
RC	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arkawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Konno S., Yamanaka I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochua H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schimi L.M., Steubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojuno M., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher R., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection.",  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK020610; BAB32146.1; -;  
 DR MGD: MGI:1925799; 9530055J05R1K.  
 SO SEQUENCE 116 AA; 13558 MW; F4E7E9AE6306A2A CRC64;

Query Match 49.0%; Score 48; DB 11; Length 116;  
 Best Local Similarity 41.2%; Pred. No. 3.1;  
 Matches 7; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 PAMMAESFGAIAHGCL 18  
 DB 36 PSWFDEAKFGVFMHGV 52

RESULT 9  
 ID 09DD22 PRELIMINARY; PRT; 452 AA.  
 AC 09DD22;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE 0610006A03RIK PROTEIN.  
 GN 0610006A03RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa Y., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,  
 RA Schiml L.M., Stahli F., Suzuki R., Tomita M., Wagner L., Mashio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection.",  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK002230; BAB21949.1; -;  
 DR MGD: MGI:1918915; 0610006A03RIK.  
 DR InterPro: IPR000933; Alpha\_L\_Fucos.  
 DR Pfam: PF01120; Alpha\_L\_Fucos; 1.  
 DR PRINTS: PR00741; GLHYDLASE29.  
 DR PROSITE: PS00385; ALPHA\_L\_FUCOSIDASE; 1.  
 SO SEQUENCE 452 AA; 52314 MW; 4BB919AA4A152B2C CRC64;

Query Match 49.0%; Score 48; DB 11; Length 452;  
 Best Local Similarity 41.2%; Pred. No. 15;  
 Matches 7; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

DB 36 PSWFDEAKFGVFMHGV 52  
 RESULT 10  
 ID 099LJ1 PRELIMINARY; PRT; 452 AA.  
 AC 099LJ1;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE RIKEN CDNA 0610006A03 GENE.  
 GN 0610006A03RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD. GROSS  
 RC TISSUE:;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC003235; AAH03235.1; -;  
 DR MGD: MGI:1918915; 0610006A03RIK.  
 DR InterPro: IPR000933; Alpha\_L\_Fucos.  
 DR Pfam: PF01120; Alpha\_L\_Fucos; 1.  
 DR PRINTS: PR00741; GLHYDLASE29.  
 DR PROSITE: PS00385; ALPHA\_L\_FUCOSIDASE; 1.  
 SO SEQUENCE 452 AA; 52280 MW; 41B319A040152B2C CRC64;

Query Match 49.0%; Score 48; DB 11; Length 452;  
 Best Local Similarity 41.2%; Pred. No. 15;  
 Matches 7; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 PAMMAESFGAIAHGCL 18  
 DB 36 PSWFDEAKFGVFMHGV 52  
 RESULT 11  
 ID 042728 PRELIMINARY; PRT; 572 AA.  
 AC 042728;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CEPHALOSPORIN ESTERASE.  
 OS Rhodospirillum rubrum (Yeast) (Rhodotorula gracilis).  
 OC Eukaryota; Fungi; Basidiomycota; Urediniomycetes;  
 OC Microbotryomycetidae; Heterogastriales; Sporidiobolaceae;  
 OC Rhodospirillum.  
 OX NCB1\_TaxID=5286;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=98069473; PubMed=9406399;  
 RX Politano M., Tonzi S.M., Burnett W.V., Romanick G., Usher J.J.;  
 RT "Purification and characterization of a cephalosporin esterase from  
 RT Rhodospirillum rubrum.",  
 RL Appl. Environ. Microbiol. 63:4807-4811(1997).  
 CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.  
 DR EMBL: AF025410; AAB93483.1; -;  
 DR HSP: P37967; 10E3.  
 DR InterPro: IPR002018; Carboxylesterase\_B.  
 DR InterPro: IPR000379; Est\_Lip\_thioestl\_actsite.  
 DR Pfam: PF00135; Coesterase; 1.  
 DR PROSITE: PS00122; CARBOXYLESTERASE\_B; 1.  
 DR KW Hydroxylase; Porin.  
 SO SEQUENCE 572 AA; 61315 MW; CC104DOC7682997E CRC64;

Query Match 49.0%; Score 48; DB 3; Length 572;

Best Local Similarity 52.9%; Pred. No. 19;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 PAMDAESFGAIAHGCL 18  
DB 442 PTYTAEGFSSAHKGL 458

RESULT 12

Q9PHM7 PRELIMINARY; PRT; 583 AA.  
AC Q9PHM7;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE ASPARTYL-TRNA SYNTHETASE (EC 6.1.1.12).  
GN ASPS OR C00640.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
OC Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIT 11168;  
RX MEDLINE=20150912; PubMed=1068204;  
RA Parish J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,  
RA Jagels K., Kariyeh A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Rajadream M.A., Rutherford K.M., Van Vleet A.H.M.,  
RA Whitehead S., Barrett B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
RT reveals hypervariable sequences.";  
RL Nature 403:665-668(2000).  
CC -1- CATALYTIC ACTIVITY: ATP + L-AMINO ACID + TRNA(AMINO ACID) = AMP +  
CC PYROPHOSPHATE + L-AMINOACYL-TRNA(AMINO ACID).  
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE (OR L-ASPARAGINE) +  
CC TRNA(ASN) = AMP + PYROPHOSPHATE + L-ASPARTYL-TRNA(ASP) (OR L-  
CC ASPARAGINYL-TRNA(ASN)).  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC -1- SIMILARITY: BELONGS TO ASPARTYL-TRNA SYNTHETASE FAMILY.  
DR EMBL; AL139075; CAB75276.1; -.  
DR HSSP; P21889; IEOR.  
DR InterPro; IPR002106; AA\_TRNA\_Ligase\_II.  
DR InterPro; IPR002312; trna-synt\_ase.  
DR InterPro; IPR002313; trna-synt\_lys\_2.  
DR InterPro; IPR004363; trna\_antl.  
DR Pfam; PF01336; trna\_antl; 1.  
DR PRINTS; PR01042; TRNASYNTHASP.  
DR PRINTS; PR00982; TRNASYNTHLVS.  
DR PROSITE; PS00179; AA\_TRNA\_Ligase\_II\_1; 1.  
DR PROSITE; PS00339; AA\_TRNA\_Ligase\_II\_2; UNKNOWN\_1.  
DR ATP-binding; Aminoacyl-trna synthetase; Complete proteome; Ligase;  
KW Protein biosynthesis.  
SQ SEQUENCE 583 AA; 66166 MW; A3723A2E100222E2 CRC64;

Query Match 48.0%; Score 47; DB 16; Length 583;  
Best Local Similarity 64.3%; Pred. No. 29;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 MDAESFGAIAHGCL 18  
DB 513 LDALSGAPPHGCI 526

RESULT 13

Q9P133 PRELIMINARY; PRT; 341 AA.  
AC Q9P133;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE PRO3091 (NEIRIN 4).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RA Zhang C., Yu Y., Zhang S., Wei H., Zhang Y., Zhou G., Bi J., Liu M.,  
RA He F.;  
RT "Functional prediction of the coding sequences of 79 new genes deduced  
RT by analysis of cDNA clones from human fetal liver."  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kawabata A., Hiki J., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Ota T., Suzuki Y., Ohashi M., Nishi T., Shibahara T.,  
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
RT "NEDO human cDNA sequencing project."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG CARCINOMA;  
RA Strausberg R.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF119916; AAF69670.1; -.  
DR EMBL; AK024691; BAB14964.1; -.  
DR EMBL; BC013591; AAH13591.1; -.  
DR HSSP; P02468; ITLE.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002049; laminin\_EGF.  
DR InterPro; IPR001134; Nectin\_C.  
DR Pfam; PF00053; laminin\_EGF; 2.  
DR Pfam; PF01759; NTR; 1.  
DR PRINTS; PR00011; EGFAMININ.  
DR SMART; SM00180; EGF\_lam; 2.  
DR SMART; SM00001; EGF\_like; 1.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_2.  
DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; UNKNOWN\_2.  
SQ SEQUENCE 341 AA; 37907 MW; 13E60C3ABBFC87 CRC64;

Query Match 46.9%; Score 46; DB 4; Length 341;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 WMDAESFGAIAHGCL 17  
DB 204 WMDAEGFSAIHGSG 217

RESULT 14

Q9NT44 PRELIMINARY; PRT; 400 AA.  
AC Q9NT44;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE HYPOPHYSICAL 44.6 KDA PROTEIN (FRAGMENT).  
GN DKEF2P43401519.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RA Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.,  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL137540; CAB70800.1; -.  
DR HSSP; P02468; ITLE.  
DR InterPro; IPR000561; EGF-like.

DR InterPro: IPR002049; Laminin\_EGF.  
 DR InterPro: IPR00134; Netrin\_C.  
 DR Pfam: PF00053; Laminin\_EGF; 3.  
 DR Pfam: PF01759; NTR; 1.  
 DR PRINTS: PR00011; EGFLAMININ.  
 DR SMART: SM00180; EGF\_Lam; 3.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_2.  
 DR PROSITE: PS01248; LAMININ\_TYPE\_EGF; UNKNOWN\_2.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 400 AA; 44649 MW; 985A7EA5BD097E01 CRC64;

Query Match 46.9%; Score 46; DB 4; Length 400;  
 Best Local Similarity 50.0%; Pred. No. 27;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 WMDAESFGAIAHGG 17  
 Db 263 WEDAQGFSAALLHSG 276

RESULT 15  
 Q9HB63 PRELIMINARY; PRT; 628 AA.  
 AC Q9HB63;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE BETA-NETRIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-20493277; PubMed-11038171;  
 RA Koch M., Murrell J.R., Hunter D.D., Olson P.F., Jin W., Keene D.R.,  
 Brunken W.J., Burgeson R.E.;  
 RT "A novel member of the netrin family, beta-netrin, shares homology  
 with the beta chain of laminin. Identification, expression, and  
 functional characterization."  
 RT J. Cell Biol. 151:221-234(2000).  
 RL EMBL: AF278532; AAG30822.1; -.  
 DR HSSP: P02468; ITLE.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR002049; Laminin\_EGF.  
 DR InterPro: IPR001886; LamNT.  
 DR InterPro: IPR00134; Netrin\_C.  
 DR Pfam: PF00053; Laminin\_EGF; 3.  
 DR Pfam: PF00055; Laminin\_Nterm; 1.  
 DR Pfam: PF01759; NTR; 1.  
 DR PRINTS: PR00011; EGFLAMININ.  
 DR ProDom: PD002082; LamNT; 1.  
 DR SMART: SM00180; EGF\_Lam; 4.  
 DR SMART: SM00136; LamNT; 1.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 SQ SEQUENCE 628 AA; 65981 MW; ABB9DBDE7EAD659 CRC64;

Query Match 46.9%; Score 46; DB 4; Length 628;  
 Best Local Similarity 50.0%; Pred. No. 45;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 WMDAESFGAIAHGG 17  
 Db 491 WEDAQGFSAALLHSG 504

Search completed: July 8, 2002, 08:21:03  
 Job time: 218 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 8, 2002, 08:18:20 ; Search time 13.48 Seconds

(without alignments)  
51.703 Million cell updates/sec

Title: US-09-582-296-1  
Perfect score: 98  
Sequence: 1 APAMDAESFGAIHGGI 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	50	51.0	461 1	P04066 homo sapien
2	49	50.0	407 1	P43713 caenorhabdi
3	46	45.9	578 1	D5BD_PSEBP
4	45	45.9	462 1	FUCO_RAT
5	44.5	45.4	227 1	DUB8_MOUSE
6	44	44.9	465 1	FUCO_CANPA
7	44	44.9	545 1	SYN_CAEEL
8	43	43.9	326 1	DUB6_HUMAN
9	42.5	43.4	807 1	YNF_PSEAE
10	42	42.9	461 1	FUCO_DICDI
11	42	42.9	591 1	SYD_PSEAE
12	41	41.8	448 1	SYN_STRPY
13	40	40.8	97 1	VGIC_BPPZA
14	40	40.8	160 1	BCHF_RHOSH
15	40	40.8	171 1	BCHF_RHOSH
16	40	40.8	173 1	LMIP_BOVIN
17	40	40.8	185 1	ISPZ_HAEIN
18	40	40.8	256 1	YAFV_ECOLI
19	40	40.8	317 1	OZEL_HUMAN
20	40	40.8	447 1	SYN_STRPY
21	40	40.8	587 1	REC_NMYCU
22	40	40.8	589 1	SYD_MYCLE
23	40	40.8	1515 1	GLTB_AZOB
24	39.5	40.3	209 1	ENMD_ECOLI
25	39.5	40.3	294 1	DAPA_CAUCR
26	39.5	40.3	305 1	PSBB_MARO
27	39.5	40.3	1335 1	PSBB_MARO
28	39	39.8	146 1	IXB_TRIFL
29	39	39.8	342 1	DUB6_MOUSE
30	39	39.8	342 1	EFTS_LACIA
31	39	39.8	345 1	EFTS_STRPN
32	39	39.8	346 1	Y193_HUMAN
33	39	39.8	389 1	VMSA_HPBVA

## ALIGNMENTS

RESULT 1	ID	FUCO_HUMAN	STANDARD:	PRT:	461 AA.
AC	P04066	Q14334	Q14335		
DT	01-NOV-1986	(Rel. 03, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Tissue alpha-L-fucosidase precursor (EC 3.2.1.51) (Alpha-L-fucosidase I) (Alpha-L-fucoside fucosyltransferase).				
GN	FUCAL				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=90026416; PubMed=2803312;				
RA	O'Connors T., Beckmann K.R., Morris C.P., Hopwood J.J.;				
RT	"Human alpha-L-fucosidase: complete coding sequence from cDNA clones.";				
RL	Biochem. Biophys. Res. Commun. 164:439-445(1989).				
RN	[2]				
RP	SEQUENCE OF 70-421 FROM N.A.				
RX	MEDLINE=88151864; PubMed=2894306;				
RA	O'Brien J.S., Williams P.J., Fukushima H., de Wet J.R., Darby J.K.,				
RT	DiGiocio R., Fowler M.L., Shows T.B.;				
RL	"Molecular biology of the alpha-L-fucosidase gene and fucosidosis.";				
RN	Enzyme 38:45-53(1987).				
RP	SEQUENCE OF 62-462 FROM N.A.				
RX	MEDLINE=85140268; PubMed=2983333;				
RA	Fukushima H., de Wet J.R., O'Brien J.S.;				
RT	"Molecular cloning of a cDNA for human alpha-L-fucosidase.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 82:1262-1265(1985).				
RN	[4]				
RP	SEQUENCE OF 70-422 FROM N.A.				
RX	MEDLINE=85076174; PubMed=6096099;				
RA	de Wet J.R., Fukushima H., Dewji N.N., Wilcox E., O'Brien J.S.,				
RT	Helinski D.R.;				
RL	"Chromogenic immunodetection of human serum albumin and alpha-L-fucosidase clones in a human hepatoma cDNA expression library.";				
RN	DNA 3:437-447(1984).				
RP	[5]				
RA	STRUCTURE OF CARBOHYDRATES.				
RT	Beem E.P., Lismann J.J.W., van Steijn G.J., van der Wal C.J.,				
RL	Trippelwitz L.A.W., Overdijk B., van Halbeek H., Mutsaers J.H.G.M.,				
RA	Vliegelandt J.F.G.;				
RT	"Structural analysis of the carbohydrate moieties of alpha-L-fucosidase from human liver.";				
RL	Glycoconj. J. 4:33-42(1987).				
RP	[6]				
RA	VARIANT ARG-281.				
RX	MEDLINE=94002224; PubMed=8399358;				
RA	Yang M., Allen H., DiGiocio R.A.;				
RT	"Pedigree analysis of alpha-L-fucosidase gene mutations in a fucosidosis family.";				

34	39	39.8	397 1	MPA2_AMBAR	P2762 ambrosia ar
35	39	39.8	481 1	GLNA_HELPY	P9485 helicobacte
36	39	39.8	1051 1	UBA2_WHEAT	P20973 triticum ae
37	39	39.8	1051 1	UBA2_WHEAT	P31251 triticum ae
38	39	39.8	2194 1	GLSN_MEDSA	003460 medicago sa
39	38.5	39.3	430 1	SYN_BACDH	09Kc78 bacillus ba
40	38.5	39.3	442 1	STD_ABRPE	03Y947 aeropyrum p
41	38.5	39.3	527 1	PUR1_MYCTU	006626 mycobacteri
42	38.5	39.3	556 1	PUR1_MYCTU	050028 mycobacteri
43	38.5	39.3	636 1	FTHS_SPIOL	P28723 spiniacia ol
44	38	38.8	278 1	YKF5_YEAST	P35731 saccharomyc
45	38	38.8	389 1	VMSA_HPBVA	P03138 hepatitis b

RL Blochim Biophys. Acta 1182:245-249(1993).  
 RN [17]  
 RP VARIANT FUCOSIDOSIS ASP-60. AND VARIANT ARG-281.  
 RX MEDLINE-93278392; PubMed-8504303.  
 RA Seo H.-C., Williams P.J., Kreitz K.A., Martin B.M., O'Brien J.S.;  
 RT "Fucosidosis: four new mutations and a new polymorphism.";  
 RL Hum. Mol. Genet. 2:423-429(1993).  
 RN [18]  
 RP VARIANT FUCOSIDOSIS LEU-63.  
 RX MEDLINE-95179129; PubMed-7874128;  
 RA Seo H.-C., Yang M., Tonlorenzi R., Williams P.J., Kim A.H.,  
 RT Fillecoame M., Gatti R., DiGioco R.A., O'Brien J.S.;  
 RT "A missense mutation (S63I) in alpha-L-fucosidase is responsible for  
 RL fucosidosis in an Italian patient.";  
 RL Hum. Mol. Genet. 3:2065-2066(1994).  
 RN [19]  
 RP VARIANT ARG-281.  
 RX MEDLINE-95114998; PubMed-7815431;  
 RA Cregg H., Winchester B., Seo H.C., O'Brien J., Swallow D.;  
 RT "Molecular basis of the common electrophoretic polymorphism (Ful/Fu2)  
 RL in human alpha-L-fucosidase.";  
 RT J. Med. Genet. 31:659-660(1994).  
 CC -1- FUNCTION: ALPHA-L-FUCOSIDASE IS RESPONSIBLE FOR HYDROLYZING THE  
 CC ALPHA-1,6-LINKED FUCOSE JOINED TO THE REDUCING-END  
 CC N-ACETYLGLUCOSAMINE OF THE CARBOHYDRATE MOETIES OF GLYCOPROTEINS.  
 CC -1- CATALYTIC ACTIVITY: An alpha-L-fucoside + H(2)O = an alcohol + L-  
 CC fucose.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SUBCELLULAR LOCATION: Lysosomal.  
 CC -1- POLYMORPHISM: THE VARIATION IN POSITION 281 IS RESPONSIBLE FOR  
 CC THE FUL/FU2 POLYMORPHISM.  
 CC -1- DISEASE: A DEFICIENCY OF THIS ACID HYDROLASE IN HUMANS AND DOGS  
 CC RESULTS IN THE LYSOSOMAL STORAGE DISEASE FUCOSIDOSIS.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 29 OF GLYCOSYL HYDROLASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR EMBL: M29877; AAA35519.1; -.  
 DR EMBL: X01390; CAA25646.1; -.  
 DR EMBL: M10355; AAA52482.1; -.  
 DR EMBL: M80815; AAA52481.1; -.  
 DR EMBL: M80810; AAA52481.1; JOINED.  
 DR EMBL: M80810; AAA52481.1; JOINED.  
 DR EMBL: M80811; AAA52481.1; JOINED.  
 DR EMBL: M80812; AAA52481.1; JOINED.  
 DR EMBL: M80813; AAA52481.1; JOINED.  
 DR EMBL: M80814; AAA52481.1; JOINED.  
 DR PIR: A33427; HHUHPA.  
 DR InterPro: IPR000933; Alpha\_L\_fucos.  
 DR Pfam: PF01120; Alpha\_L\_fucos. 1.  
 DR PRINTS: PR00741; GLHYDRASE29.  
 DR PROSITE: PS00385; ALPHA\_L\_FUCOSIDASE. 1.  
 DR Hydrolase; Glycosidase; Lysosome; Glycoprotein; Signal; Polymorphism;  
 KW Disease mutation.  
 KM  
 FT SIGNAL 1 22  
 FT CHAIN 23 461  
 FT ACT\_SITE 291 291  
 FT CARBOHYD 236 236  
 FT CARBOHYD 263 263  
 FT CARBOHYD 377 377  
 FT VARIANT 60 60  
 FT  
 FT TISSUE ALPHA-L-FUCOSIDASE.  
 FT MAY BE IMPORTANT FOR CATALYSIS.  
 FT N-LINKED (GLCNAC . . . ) (POTENTIAL).  
 FT N-LINKED (GLCNAC . . . ) (POTENTIAL).  
 FT G -> D (IN FUCOSIDOSIS; LOSS OF  
 FT ACTIVITY).  
 FT /FTID=VAR\_002442.  
 FT S -> L (IN FUCOSIDOSIS).  
 FT /FTID=VAR\_002443.  
 FT  
 FT VARIANT 63 63  
 FT VARIANT 281 281  
 FT O -> R.

FT	CONFLICT	71	72	/FTID=VAR_002444.
FT	CONFLICT	264	264	WF -> FL (IN REF. 3 AND 4).
FT	CONFLICT	422	422	C -> S (IN REF. 1 AND 2).
FT	CONFLICT	461	53176	Q -> P (IN REF. 4).
SO	SEQUENCE	461 AA;	53176 MW;	EAA4AF7119C113CB CRC64;

QY	Query Match	51.0%;	Score 50;	DB 1;	Length 461;
QY	Best Local Similarity	47.1%;	Pred. No. 1.3;	Mismatches 8;	Indels 0;
Db	Matches 8;	Conservative 1;	Mismatches 8;	Indels 0;	Gaps 0;
	2	PAMMDAESFGAIAHGGL 18			
	45	PAMFDEAKFGVFIHMGV 61			

RESULT	2
ID	FUCO_CAEEL
AC	P49713
DT	01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	01-OCT-1996 (Rel. 34, Last annotation update)
DE	Purative alpha-L-fucosidase precursor (EC 3.2.1.51) (Alpha-L-fucoside fucosyltransferase).
DE	W03G11.3.
OS	Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC	Rhabditidae; Peloderinae; Caenorhabditis.
OX	NCBI_Taxid=6239;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BRISTOL NZ;
RA	McMurray A.;
RL	Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC	-I FUNCTION: ALPHA-L-FUCOSIDASE IS RESPONSIBLE FOR HYDROLYZING THE
CC	ALPHA-1,6-LINKED FUCOSE JOINED TO THE REDUCING-END
CC	N-ACETYLGLUCOSAMINE OF THE CARBOHYDRATE MOIETIES OF GLYCOPROTEINS
CC	(BY SIMILARITY)
CC	-I CATALYTIC ACTIVITY: An alpha-L-fucoside + H(2)O = an alcohol + L-
CC	fucose.
CC	-I SIMILARITY: BELONGS TO FAMILY 29 OF GLYCOSYL HYDROLASES.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	at the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; 267738; CAA91546.1; -
DR	WoimPep; W03G11.3; CE03783.
DR	InterPro; IPR000933; Alpha_L_Fucos.
DR	Pfam; PF01120; Alpha_L_fucos; 1.
DR	PRINTS; PR00741; GLHYDRLASE29.
DR	PROSITE; PS00385; ALPHA_L_FUCOSIDASE; 1.
KW	Hypothetical protein; Hydrolase; Glycosidase; Signal.
FT	SIGNAL 1 16
FT	CHAIN 17 407
FT	ACT_SITE 291 291
FT	POTENTIAL.
FT	POTENTIAL ALPHA-L-FUCOSIDASE.
FT	MAY BE IMPORTANT FOR CATALYSIS (BY
FT	SIMILARITY).
FT	CARBOHYD 197 197
FT	CARBOHYD 358 358
FT	CARBOHYD 374 374
SO	SEQUENCE 407 AA; 47999 MW; F28652D614347E20 CRC64;

QY	Query Match	50.0%;	Score 49;	DB 1;	Length 407;
QY	Best Local Similarity	47.1%; <th>Pred. No. 1.7;</th> <td></td> <td></td>	Pred. No. 1.7;		
Matches	8;	Conservative 1;	Mismatches 8;	Indels 0;	Gaps 0;
	2	PAMMDAESFGAIAHGGL 18			

Db 31 PSWYDSKFGFCHWGL 47

RESULT 3  
DSBD\_PSESP STANDARD: PRT: 578 AA.

AC Q9KJ23;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Thiol:disulfide interchange protein dsb precursor.  
GN DSBD.

OS Pseudomonas sp.  
OC Bacteria; Proteobacteria.  
OX NCBI\_TaxId=306;  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JRI / K1;

RA Johann A., Michel J., Averhoff B., Gottschalk G.;

RT "The two-component signal transduction system Arms is involved in  
regulation of 3-isopropylcatechol dioxygenase in the mutant strain  
Pseudomonas sp. K1."

RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

CC DISULFIDE BONDS IN SOME PERIPLASMIC PROTEINS AND FOR THE ASSEMBLY  
OF THE PERIPLASMIC C-TYPE CYTOCHROMES. ACTS BY TRANSFERRING  
ELECTRONS FROM CYTOPLASMIC THIOREDOXIN TO THE PERIPLASM. THIS  
TRANSFER INVOLVES A CASCADE OF DISULFIDE BOND FORMATION AND  
REDUCTION STEPS (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
(By similarity).

CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY. DSBD SUBFAMILY.  
(By similarity).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AF155506; AAF80267.1; -  
DR InterPro: IPR000063; Thioredoxin.  
KW Redox-active center; Electron transport; Transmembrane;  
Inner membrane; Cytochrome c-type biogenesis; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 578  
FT DOMAIN 19 158  
FT TRANSMEM 159 179  
FT DOMAIN 180 207  
FT TRANSMEM 208 228  
FT TRANSMEM 229 237  
FT TRANSMEM 238 258  
FT TRANSMEM 259 281  
FT TRANSMEM 282 302  
FT TRANSMEM 303 321  
FT TRANSMEM 322 342  
FT TRANSMEM 343 351  
FT TRANSMEM 352 372  
FT TRANSMEM 373 373  
FT TRANSMEM 374 394  
FT TRANSMEM 395 408  
FT TRANSMEM 409 429  
FT TRANSMEM 430 578  
FT DISULFID 118 124  
FT DISULFID 179 229  
FT DISULFID 491 494  
FT SEQUENCE 578 AA; 62181 MW; 7473FC17A9BF7284 CRC64;

Query Match

46.9%; Score 46; DB 1; Length 578;

Best Local Similarity 53.8%; Pred. No. 7.3;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 AAMDASFGLAH 15  
Db 396 AMKSEGFRAH 408

RESULT 4

FUCO\_RAT STANDARD: PRT: 462 AA.

AC PI7164;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Tissue alpha-L-fucosidase precursor (EC 3.2.1.51) (Alpha-L-fucosidase  
I) (Alpha-L-fucoside fucohydrolase).

GN FUCAL.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxId=10116;

RL [1]  
SEQUENCE FROM N.A., AND SEQUENCE OF 29-40; 90-124 AND 307-372.

CC TISSUE=Liver;  
MEDLINE=90147522; Pubmed=2482732;

RA Fisher K.J., Aronson N.N. Jr.;

RT "Isolation and sequence analysis of a cDNA encoding rat liver  
alpha-L-fucosidase."

CC BIOCHEM. J. 264:695-701(1989).

CC -1- FUNCTION: ALPHA-L-FUCOSIDASE IS RESPONSIBLE FOR HYDROLYZING THE  
ALPHA-1,6-LINKED FUCOSE JOINED TO THE REDUCING-END  
N-ACETYLGLUCOSAMINE OF THE CARBOHYDRATE MOIETIES OF GLYCOPROTEINS.

CC -1- CATALYTIC ACTIVITY: An alpha-L-fucoside + H(2)O = an alcohol + L-  
fucose.

CC -1- SUBUNIT: HOMOTETRAMER.

CC -1- SUBCELLULAR LOCATION: Lysosomal.

CC -1- SIMILARITY: BELONGS TO FAMILY 29 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X16145; CAA34268.1; -

DR PIR: S07074; S07074.

DR PIR: S10235; S10235.

DR InterPro: IPR000933; Alpha L. fucos.

DR Pfam: PR01120; Alpha L. fucos. 1.

DR PRINTS: PR00741; GHIHDLASE29.

DR PROSITE: PS00385; ALPHA-L-FUCOSIDASE; 1.

KW Hydrolase; Glycosidase; Lysosome; Glycoprotein; Signal.

FT SIGNAL 1 28

FT CHAIN 29 462

FT ACT\_SITE 292 292

FT CARBOHYD 237 237

FT CARBOHYD 264 264

FT CARBOHYD 378 378

FT SEQUENCE 462 AA; 53486 MW; B9B06B62E6019C15 CRC64;

SO

Query Match 45.9%; Score 45; DB 1; Length 462;  
Best Local Similarity 41.2%; Pred. No. 8.4;  
Matches 7; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

OY 2 PAMDAESFGAIAHGL 18  
Db 46 PAMDEAKFGIFVHWGV 62

RESULT 5  
DJB8 MOUSE STANDARD: PRT: 227 AA.  
AC 09017;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DnaJ homolog subfamily B member 8 (mdj6).  
GN DnaJb8.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD-1;  
RA MEDLINE=21023480; PubMed=11147971;  
RA Ohtsuka K., Hata M.;  
RT "Mammalian HSP40/DNAJ homologs: cloning of novel cDNAs and a proposal  
RT for their classification and nomenclature."  
RL Cell Stress Chaperones 5:98-112(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Testis;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kiehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schirml L.M., Staudl F., Suzuki R., Tomita M., Wagner I., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carrinci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bolt C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,  
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AB028856; BAA88304.1; -  
DR EMBL: AK006026; BAB24372.1; -  
DR HSSP: P08622; 1BQZ.  
DR MGI: MGI:1932801; DnaJb8.  
DR InterPro: IPR001623; DnaJb8.  
DR Pfam: PF00226; DnaJ.1.  
DR SMART: SM00271; DnaJ.1.  
DR PROSITE: PS00636; DnaJ.1; 1.  
DR PROSITE: PS50076; DnaJ\_2; 1.  
KW Chaperone.  
FT DOMAIN 3 69 J-DOMAIN.  
SQ SEQUENCE 227 AA; 25230 MW; 7A6E4CB7989FC095 CRC64;

QY 2 PAMMDA-ESFGAIHGC 17  
ID 1:1 11 :11  
DB 140 PAFMEALSSFNLTGHGC 156  
RESULT 6  
FICO\_CANFA STANDARD: PRT: 465 AA.  
ID FICO\_CANFA  
AC P48300;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Tissue alpha-L-fucosidase precursor (EC 3.2.1.51) (Alpha-L-fucosidase  
DE I) (Alpha-L-fucoside fucosyltransferase).  
GN FUCAL.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ENGLISH SPRINGER SPANIEL; TISSUE=Liver, and Blood;  
RX MEDLINE=96282899; PubMed=8730282;  
RA Skelly B.J., Sargan D.R., Heritage M.E., Winchester B.G.;  
RT "The molecular defect underlying canine fucosidosis."  
RL J. Med. Genet. 33:284-288(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ENGLISH SPRINGER SPANIEL;  
RX MEDLINE=96325097; PubMed=8661697;  
RA Ochiodoro T., Anson D.S.;  
RT Isolation of the canine alpha-L-fucosidase cDNA and definition of  
RT the fucosidosis mutation in English Springer Spaniels."  
RL Mamm. Genome 7:271-274(1996).  
CC -1- FUNCTION: ALPHA-L-FUCOSIDASE IS RESPONSIBLE FOR HYDROLYZING THE  
CC ALPHA-1,6-LINKED FUCOSE JOINED TO THE REDUCING-END  
CC N-ACETYLGLUCOSAMINE OF THE CARBOHYDRATE MOIETIES OF GLYCOPROTEINS.  
CC -1- CATALYTIC ACTIVITY: An alpha-L-fucoside + H(2)O = an alcohol + L-  
CC fucose.  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- SUBCELLULAR LOCATION: Lysosomal.  
CC -1- DISEASE: A DEFICIENCY OF THIS ACID HYDROLASE IN HUMANS AND DOGS  
CC RESULTS IN THE LYSOSOMAL STORAGE DISEASE FUCOSIDOSIS.  
CC -1- SIMILARITY: BELONGS TO FAMILY 29 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X92671; CAA63362.1; -  
DR EMBL: X92672; CAA63362.1; JOINED.  
DR EMBL: X92673; CAA63362.1; JOINED.  
DR EMBL: X92674; CAA63362.1; JOINED.  
DR EMBL: X92675; CAA63362.1; JOINED.  
DR EMBL: X92676; CAA63362.1; JOINED.  
DR EMBL: X92677; CAA63362.1; JOINED.  
DR EMBL: X92678; CAA63362.1; JOINED.  
DR EMBL: X92679; CAA63362.1; JOINED.  
DR EMBL: X92448; CAA63197.1; -  
DR EMBL: X92449; CAA63197.1; -  
DR EMBL: U29765; AAB17403.1; -  
DR EMBL: U29766; AAB17401.1; -  
DR InterPro: IPR000933; Alpha.L.fucos.  
DR Pfam: PF01120; Alpha.L.fucos.  
DR PRINTS: PR00741; GLYHDRLASE29.  
DR PROSITE: PS00385; ALPHA\_L\_FUCOSIDASE.1.  
KW Hydrolase; Glycosidase; Lysosome; Glycoprotein; Signal.  
FT SIGNAL 1 26  
FT CHAIN 27 465 TISSUE ALPHA-L-FUCOSIDASE.  
FT ACT\_SITE 256 296 MAY BE IMPORTANT FOR CATALYSIS.  
FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 14 15 AV -> LPE (IN REF. 2).  
FT CONFLICT 30 32 AAA -> GEP (IN REF. 2).  
FT CONFLICT 66 66 E -> V (IN REF. 2).  
FT CONFLICT 234 234 K -> E (IN REF. 2).  
FT CONFLICT 257 257 D -> G (IN REF. 1); CAA63197).  
FT CONFLICT 291 291 L -> H (IN REF. 2).  
SQ SEQUENCE 465 AA; 53757 MW; 832BA3C510341B7 CRC64;

Query Match 44.9%; Score 44; DB 1; Length 465;  
Best Local Similarity 43.8%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Oy 2 PAMDAESFGALHGG 17  
Db 50 PDWFDKAKFGVFWHG 65

RESULT 7  
SYN\_CAEEL STANDARD; PRT; 545 AA.  
ID SYN\_CAEEL  
AC 019722;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Probable asparaginyl-tRNA synthetase, cytoplasmic (EC 6.1.1.22)  
DE (Asparagine--tRNA ligase) (Asnrs).  
GN F22D6.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Poloderinae; Caenorhabditis.  
ON NCBI\_TaxID=6239;  
RN SEQUENCE FROM N.A.  
RP STRAIN-BRISTOL N2;  
RA Wilkinson J.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
CC -I- CATALYTIC ACTIVITY: ATP + L-asparagine + tRNA(Asn) = AMP +  
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -I- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; 271262; CAA95808.1; -  
DR HSSP; 052428; 1B8A.  
DR WormPep; F22D6.3; CE05684.  
DR InterPro; IPR002106; AA.TRNA.Ligase\_II.  
DR InterPro; IPR002309; tRNA-synt\_2.  
DR InterPro; IPR002312; tRNA-synt\_2asp.  
DR Pfam; PF00152; tRNA-synt\_2; 1.  
DR Pfam; PF01336; tRNA-anti\_1.  
DR PRINTS; PRO1042; TRNASYNTHASP.  
DR PROSITE; PS00179; AA.TRNA.LIGASE\_II.1; FALSE NEG.  
DR PROSITE; PS00339; AA.TRNA.LIGASE\_II.2; FALSE NEG.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.  
SQ SEQUENCE 545 AA; 61185 MW; 53FADA68ECB81BAC CRC64;

Query Match 44.9%; Score 44; DB 1; Length 545;  
Best Local Similarity 42.9%; Pred. No. 14;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Oy 4 WMDAESFGALHGG 17  
Db 50 PDWFDKAKFGVFWHG 65

Db 499 YMDQRKGYSPVHGG 512  
RESULT 8  
ID DB6\_HUMAN STANDARD; PRT; 326 AA.  
AC 075190; 095806; Q9UIK6;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DnaJ homolog subfamily B member 6 (heat shock protein J2) (HSJ-2)  
DE (MSJ-1) (HHDJ1) (MRJ).  
GN DNAJB6 OR HSJ2 OR MSJ1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
ON NCBI\_TaxID=9606;  
RN SEQUENCE FROM N.A. (ISOFORM B).  
RP TISSUE-Testis;  
RA Saito T., Seki N.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN  
RN SEQUENCE FROM N.A. (ISOFORM B).  
RP TISSUE-Testis;  
RA "A new member of human dnaJ-related gene family."  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN  
RN SEQUENCE FROM N.A. (ISOFORM B).  
RP TISSUE-Testis;  
RA "HSJ2, a novel human homologue of the bacterial heat-shock protein  
RT DnaJ."  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN  
RN SEQUENCE FROM N.A. (ISOFORM B).  
RP TISSUE-Testis;  
RA "Pituitary tumor-transforming gene protein associates with ribosomal  
RT protein S10 and a novel human homologue of DnaJ in testicular cells."  
RL J. Biol. Chem. 274:3151-3158(1999).  
RN  
RN SEQUENCE FROM N.A. (ISOFORMS AND B).  
RA Hanai R.;  
RT "DnaJ homologue HSJ2b."  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN  
RN SEQUENCE FROM N.A. (ISOFORM B).  
RA Zhang J.S., Nelson M., Wang L., Smith D.I.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
RN  
RN SEQUENCE FROM N.A. (ISOFORM B).  
RP TISSUE-Kidney;  
RC MEDLINE=21154917; PubMed=11230166;  
RA Wiemann S., Weill B., Wellenreuther R., Gassenhuber J., Gassl S.,  
RA Ansgore W., Boecher M., Bloeker H., Bauersachs S., Blum H.,  
RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,  
RA Lewes H.-W., Ottenweider B., Oberwalter B., Tampe J., Heubner D.,  
RA Wamplit R., Korn B., Klein M., Poustka A.;  
RT "Towards a catalog of human genes and proteins: sequencing and  
RT analysis of 500 novel complete protein coding human cDNAs."  
RL Genome Res. 11:422-435(2001).  
RN  
RN SEQUENCE FROM N.A. (ISOFORMS AND B).  
RP TISSUE-Skin, and Placenta;  
RC Strausberg R.;  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
CC -I- SUBUNIT: INTERACTS WITH PTG.  
CC -I- ALTERNATIVE PRODUCTS: 2 isoforms; a (shown here) and b, are  
CC produced by alternative splicing.  
CC -I- SIMILARITY: CONTAINS 1 J DOMAIN.  
CC -I- CAUTION: Ref.3 sequence differs from that shown due to a  
CC frameshift in position 197.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

**RT, "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome**

RT "A developmentally regulated gene product from Dictyostelium  
RA Mueller-Taubenberger A., Westphal M., Noegel A., Gerisch G.;  
RA Mueller-Taubenberger A., Westphal M., Noegel A., Gerisch G.;

RT "A developmentally regulated gene product from Dictyostelium  
RA Mueller-Taubenberger A., Westphal M., Noegel A., Gerisch G.;  
RA Mueller-Taubenberger A., Westphal M., Noegel A., Gerisch G.;

RT discoidium shows high homology to human alpha-L-fucosidase.;  
RL FEBS Lett. 246:185-192(1989).  
CC -1- FUNCTION: ALPHA-L-FUCOSIDASE IS RESPONSIBLE FOR HYDROLYZING THE  
CC ALPHA-1,6-LINKED FUCOSE JOINED TO THE REDUCING-END  
CC N-ACETYLGLUCOSAMINE OF THE CARBOHYDRATE MOIETIES OF GLYCOPROTEINS.  
CC -1- CATALYTIC ACTIVITY: An alpha-L-fucoside + H<sub>2</sub>O = an alcohol + L-  
CC fucose.  
CC MISCELLANEOUS: THE MAXIMUM EXPRESSION OF ALPHA-L-FUCOSIDASE OCCURS  
CC DURING AGGREGATION STAGE OF DICTYOSTELIUM DISCOIDEUM.  
CC -1- SIMILARITY: BELONGS TO FAMILY 29 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Y07497; CAA68800.1; -  
DR PIR: A30364; A30364.  
DR DictyDb; DD02008; alfa.  
DR InterPro; IPR000933; Alpha\_L\_fucos.  
DR Pfam; PF01120; Alpha\_L\_fucos; 1.  
DR PRINTS; PR00741; GLHYDRLASE29.  
DR PROSITE; PS00385; ALPHA\_L\_FUCOSIDASE; 1.  
DR Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 18 POTENTIAL.  
FT CHAIN 19 461 ALPHA-L-FUCOSIDASE.  
FT AC\_SITE 288 288 MAY BE IMPORTANT FOR CATALYSIS.  
SQ SEQUENCE 461 AA; 52647 MW; CB3D0771FFAB516 CRC64;  
  
Query Match 42.9%; Score 42; DB 1; Length 461;  
Best Local Similarity 41.2%; Pred. No. 25;  
Matches 7; Conservative 1; Mismatches 9; Indels 0; Gaps 0;  
  
QY 2 PAMMDSEFGAIAHGGL 18  
Db 34 PGWYDVVAFGFIHFGI 50

RA Hishida T., Iwasaki H., Ishioke K., Shinagawa H.;  
RT "Molecular analysis of the Pseudomonas aeruginosa genes, ruvA, ruvB  
RT and ruvC, involved in processing of homologous recombination  
RT intermediates";  
RL Gene 182:63-70(1996).  
CC -1- CATALYTIC ACTIVITY: ATP + L-aspartate + tRNA(Asp) = AMP +  
CC diphosphate + L-asparaginyl-tRNA(Asp).  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AE004530; AAG04352.1; -  
DR DB: D83138; BAA11815.1; -  
DR HSPR; P21889; 1E0R.  
DR InterPro; IPR002106; AA\_tRNA\_ligase\_II.  
DR InterPro; IPR004115; GAD.  
DR InterPro; IPR002309; tRNA-synt\_2.  
DR InterPro; IPR002312; tRNA-synt\_asp.  
DR Pfam; PF02938; GAD; 1.  
DR Pfam; PF00152; tRNA-synt\_2; 2.  
DR Pfam; PF01356; tRNA-anti\_1.  
DR PRINTS; PR01042; TRNASYNTHASP.  
DR PROSITE; PS00179; AA\_tRNA\_LIGASE\_II\_1; 1.  
DR PROSITE; PS00339; AA\_tRNA\_LIGASE\_II\_2; 1.  
KM Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;  
FT COMPLETE 588 P -> A (IN REF. 2).  
FT CONFLICT 588  
SQ SEQUENCE 591 AA; 66207 MW; 62B278CA0DEDF0A4 CRC64;  
  
Query Match 42.9%; Score 42; DB 1; Length 591;  
Best Local Similarity 57.1%; Pred. No. 32;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
QY 5 MDAESFGAIAHGGL 18  
Db 519 LDALIKYCAPPHGGL 532

RESULT 11  
SYD\_PSEAE  
ID SYD\_PSEAE STANDARD; PRT; 591 AA.  
AC 051422;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Asparaginyl-tRNA synthetase (EC 6.1.1.22) (Asparagine--tRNA ligase)  
DE (AspRS).  
GN ASPS OR PA0963.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=2043737; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Gabor R.L., Goltry L., Tolerinto E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laibig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
RN [2]  
RP SEQUENCE OF 231-591 FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=97136691; PubMed=8982068;

RESULT 12  
SYN\_STRPY  
ID SYN\_STRPY STANDARD; PRT; 448 AA.  
AC 09A0R9;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Asparaginyl-tRNA synthetase (EC 6.1.1.22) (Asparagine--tRNA ligase)  
DE (AspRS).  
GN ASNS OR SPY0651.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;  
RX MEDLINE=21192684; PubMed=11296296;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
CC -1- CATALYTIC ACTIVITY: ATP + L-asparagine + tRNA(Asn) = AMP +  
CC diphosphate + L-asparaginyl-tRNA(Asn).

CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AE006519; AKK3614.1; -.  
CC HSSP: 052428; 188A.  
CC InterPro: IPR002106; ALCRNA\_LigaseII.  
CC InterPro: IPR002312; tRNA-synt-asp.  
CC InterPro: IPR004365; tRNA-anti.  
CC Pfam: PF01336; tRNA-anti.1.  
CC PRINTS: PR01042; TRNASYNTHASP.  
CC PROSITE: PS00179; AA-TRNA\_LIGASE\_IL1\_1.  
CC PROSITE: PS00339; AA-TRNA\_LIGASE\_IL2\_1.  
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
CC Complete proteome.  
SQ SEQUENCE 448 AA; 51207 MW; E19800F0F4F61E97 CRC64;

Query Match 41.8%; Score 41; DB 1; Length 448;  
Best Local Similarity 35.7%; Pred. No. 36;  
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 4 WMDAESFGAIAHGC 17  
Db 402 YIDLKRYGSPVPHG 415

RESULT 13  
VGIC\_BPPZA  
ID VGIC\_BPPZA STANDARD; PRT; 97 AA.  
AC P06949;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 01-JAN-1988 (Rel. 06, Last annotation update)  
DE Early protein GPIC.  
GN IC.  
OS Bacteriophage PZA.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae.  
OX NCBI\_TaxID=10757;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86056991; PubMed=3934048;  
RA Paces V., Vlcek C., Urbanek P., Hostomsky Z.;  
RT "Nucleotide sequence of the major early region of Bacillus subtilis  
RT phage PZA, a close relative of phi 29."  
RL Gene 38:45-56(1985).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M1813; AAA8479.1; -.  
CC PIR: C24528; ERBP12.  
DR EMBL: M1813; AAA8479.1; -.  
DR PIR: C24528; ERBP12.  
KW Early protein.  
SQ SEQUENCE 97 AA; 11159 MW; 40FECB8D399FA25 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 97;  
Best Local Similarity 50.0%; Pred. No. 12;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 AWMDAESFGAIAHG 16  
Db 82 AMVIVLGEVAVHG 95

RESULT 14  
BCHF\_RHOSH  
ID BCHF\_RHOSH STANDARD; PRT; 160 AA.  
AC Q53222;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 2-vinyl bacteriochlorophyllide hydratase (EC 4.2.1.-).  
GN BCHF.  
OS Rhodospirillum rubrum (Rhodospirillum rubrum).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;  
OC Rhodospirillum.  
OX NCBI\_TaxID=1063;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIB 8253;  
RA Naylor G.W., Adlesee H.A., Gibson L.C.D., Hunter C.N.;  
RT "The photosynthesis gene cluster of Rhodospirillum rubrum."  
RT Photosyn. Res. 62:121-139(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;  
RX MEDLINE=20115911; PubMed=10648776;  
RA Choudhary M., Kaplan S.;  
RT "DNA sequence analysis of the photosynthesis region of Rhodospirillum  
RT sphaeroides 2.4.1.";  
RT Nucleic Acids Res. 28:862-867(2000).  
RN [3]  
RP SEQUENCE OF 1-43 FROM N.A.  
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;  
RA Gomelsky M., Kaplan S.;  
RT "Genetic evidence that Ppsr from Rhodospirillum rubrum 2.4.1  
RT functions as a repressor of puc and bchf expression."  
RT Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- PATHWAY: Light-independent bacteriochlorophyll biosynthesis.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AJ010302; CAB38726.1; -.  
CC EMBL: AF195122; AAF24276.1; -.  
CC EMBL: L37197; AAA51373.1; -.  
CC Photosynthesis; Bacteriochlorophyll biosynthesis; Lyase.  
SQ SEQUENCE 160 AA; 17984 MW; B2DB5118BD6E24 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 160;  
Best Local Similarity 62.5%; Pred. No. 19;  
Matches 10; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

QY 1 APA--WMDAESFGAIA 14  
Db 94 APAFMEDVFSFGVIA 109

RESULT 15  
BCHF\_RHOC  
ID BCHF\_RHOC STANDARD; PRT; 171 AA.  
AC P26165;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE 2-vinyl bacteriochlorophyllide hydratase (EC 4.2.1.-).

GN BCHF.  
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
OC Rhodobacter.  
OX NCBI\_TaxID-1061;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SB1003;  
RX MEDLINE-93224465; PubMed-8385667;  
RA Burke D.H., Alberti M., Hearst J.E.;  
RT "bchNBNH bacteriochlorophyll synthesis genes of Rhodobacter  
RT capsulatus and identification of the third subunit of  
RT light-independent protochlorophyllide reductase in bacteria and  
RT plants.";  
RU J. Bacteriol. 175:2414-2422(1993)  
CC -i- PATHWAY: Light-Independent Bacteriochlorophyll biosynthesis.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z11165; CAA77527.1; -  
DR PIR: S17811; S17811.  
DR PIR: A49851; A49851.  
KW Photosynthesis; Bacteriochlorophyll biosynthesis; Lyase.  
SQ SEQUENCE 171 AA: 19254 MW: 62F5CE7AA15F3E1C CRC64;

Query Match 40.8%; Score 40; DB 1; Length 171;  
Best Local Similarity 62.5%; Pred. No. 20;  
Matches 10; Conservative 0; Mismatches 4; Indels 2; Gaps 1;  
OY 1 APA-WMDAESFGAIA 14  
||| | | |||  
DB 94 APAFWEDVFSFGVIA 109

Search completed: July 8, 2002, 08:21:22  
Job time: 182 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:10:40 ; Search time 26.89 Seconds

(without alignments)  
64,322 Million cell updates/sec

Title: US-09-582-296-1

Perfect score: 98  
Sequence: 1 APAMDAESFGALAHGSL 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : PIR71:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	51.0	461	1	HMHUFA
2	49	50.0	407	2	T26127
3	47	48.0	583	2	G81412
4	46	46.9	400	2	T46383
5	45.5	46.4	352	2	B85518
6	45.5	46.4	378	2	F90667
7	45	45.9	224	2	AG0406
8	45	45.9	363	2	A83563
9	45	45.9	384	2	H87019
10	45	45.9	449	2	G72393
11	45	45.9	462	2	S10235
12	44	44.9	517	2	D86637
13	44	44.9	545	2	T21253
14	43.5	44.4	258	2	A69830
15	43	43.9	242	2	A11883
16	42.5	43.4	769	2	B87681
17	42.5	43.4	808	2	F90915
18	42.5	43.4	808	2	F64914
19	42.5	43.4	808	2	C85764
20	42	42.9	334	2	E69361
21	42	42.9	361	2	PC4295
22	42	42.9	461	1	A30364
23	42	42.9	483	2	F75360
24	42	42.9	541	2	T48811
25	42	42.9	591	2	E83524
26	42	42.9	700	2	I51235
27	42	42.9	1486	2	F86785
28	41.5	42.3	606	2	T35378
29	41	41.8	76	2	E87562

30	41	41.8	81	2	A90488	alpha-fucosidase N
31	41	41.8	414	2	A83745	N-carbamyl-L-amino
32	41	41.8	419	2	G83355	hypothetical prote
33	41	41.8	472	2	AG3600	cellulose synthase
34	41	41.8	498	2	F82651	hypothetical prote
35	41	41.8	589	2	H82627	aspartyl-tRNA synt
36	41	41.8	644	2	C75457	alpha-amylase - De
37	41	41.8	877	2	A83437	probable ClpA/B-ty
38	41	41.8	1527	2	H70655	glutamate synthase
39	41	41.8	1527	2	E86916	hypothetical prote
40	41	41.8	1530	2	E82085	glutamate synthase
41	40.5	41.3	242	2	F87640	exsb protein (limp
42	40.5	41.3	474	2	S65763	chitinase (EC 3.2.
43	40	40.8	97	1	ERBP12	gene 1C protein -
44	40	40.8	109	2	C75409	conserved hypothet
45	40	40.8	148	2	H71007	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

HMHUFA

alpha-L-fucosidase (EC 3.2.1.51) 1 precursor, tissue - human

N:Alternate names: alpha-L-fucosidase I; alpha-L-fucoside fucosylolase

C:Species: Homo sapiens (man)

C:Date: 30-Jun-1987 #sequence.revision 30-Jun-1991 #text.change 21-Jul-2000

C:Accession: A33427; A00901; A61016; I54348; I37394; I84457

R:Biochem. Biophys. Res. Commun. 164, 439-445, 1989

```

A:Molecule type: mRNA
A:Residues: 70, 'FL', 73-263, 'C', 265-421, 'P', <RE3>
A:Cross-references: CB:M0157, MID:g182778, PID:g182779
C:Comment: The fucosidases hydrolyze alpha-L-fucose from glycolipids and oligosaccharides
C:Genetics:
A:Gene: GDB:FUCAL
A:Cross-references: GDB:119237, OMIM:230000
A:Map position: 1p35-1p34
A:Introns: 125/2; 170/2; 216/2; 251/3; 318/3; 382/2; 415/3
C:Superfamily: alpha-L-fucosidase
C:Keywords: fucosidosis; glycoprotein; glycosidase; homotetramer; hydrolase; lysosome
F:1-32/Domain: signal sequence #status predicted <SIG>
F:23-461/Product: alpha-L-fucosidase #status predicted <MAT>
F:230,246,263,377/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match          51.0%; Score 50; DB 1; Length 461;
Best Local Similarity 47.1%; Pred. No. 2.3;
Matches      8; Conservative      1; Mismatches      8; Indels      0; Gaps      0;

Oy      2 PAMMDAESFGALHGL 18
        ||| | | | | |
Db      45 PAMDEAKFGVTHMGV 61

```

```

Query Match Similarity      51.0%; Score 50; DB 1; Length 461;
Best Local Similarity      47.1%; Pred. No. 2.3;
Matches      8; Conservative      1; Mismatches      8; Indels      0; Gaps      0;

OY      2 PAMMDAESFGAIAHGGL 18
      ||| | | | | |
Db      45 PAMFDEAKFGVFIFHWGV 61

RESULT      2

T26127
Hypothetical protein W03G11.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
C:Accession: T26127
R:McMurray, A.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z20156
A:Accession: T26127
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-407 <MIL>
A:Cross-references: EMBL:Z67738; PIDN:CAA91546.1; GSPDB:GN00028; CESP:W03G11.3
A:Experimental source: clone W03G11
C:Genetics:
A:Gene: CESP:W03G11.3
A:Map position: X
A:Introns: 19/3; 49/3; 127/2; 203/3; 248/2; 355/2
C:Superfamily: alpha-L-fucosidase

Query Match      50.0%; Score 49; DB 2; Length 407;
Best Local Similarity      47.1%; Pred. No. 2.9;
Matches      8; Conservative      1; Mismatches      8; Indels      0; Gaps      0;

OY      2 PAMMDAESFGAIAHGGL 18
      ||| | | | | |
Db      31 PSWYDSKFGIFCHWGL 47

RESULT      3

G81412
aspartate--tRNA ligase (EC 6.1.1.12) Cj0640c [imported] - Campylobacter jejuni (strain NCTC 11168)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000
C:Accession: G81412
R:Packhill, J., Wren, B.W., Mungall, K., Ketley, J.M., Churcher, C., Basham, D., Chillingworth, C.W., Quail, M., Rajandream, M.A., Rutherford, K.M., VanVleet, A., Whitehead, S., Barrett, N.A., Tate, J.
Nature 403 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912
A:Accession: G81412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-583 <PAR>
A:Cross-references: GB:AL133075; GB:AL111168; NID:g6967817; PIDN:CAB75276.1; PID:g696810
A:Experimental source: serotype O2, strain NCTC 11168

```

C;Genetics:  
A;Gene: asps; Cj0640c  
C;Superfamily: lysine--tRNA ligase  
C;Keywords: ligase

Query Match	48.0%	Score 47	DB 2	Length 583
Best Local Similarity	64.3%	Pred. No. 9.5		
Matches	9	Conservative	2	Mismatches 3
				Indels 0
				Gaps 0

```
QY      5 MDAESFGAIAHGGL 18
          :||| |||:
Db     513 LDALSFQAPPHGGI 526
```

RESULT 4  
T46383  
hypothetical protein DKFZp434O1519.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T46383  
R:Ottenwaelder, B.; Obermayer, B.; Mewes, H.-W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z23031  
A:Accession: T46383  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-400 <AAAs>  
A:Cross-references: EMBL:AL137540  
A:Experimental source: adult testis, clone DKFZp434O1519  
A:Genetics:  
N:Note: DKFZp434O1519.1

Query Match	46.9%	Score 46	DB 2	length 400
Best Local Similarity	50.0%	Pred. No. 9		
Matches	7	Conservative	2	Mismatches 5; Indels 0; Gaps 0;
OY	4	WMDAESFGAIIHGG	17	
Db	263	WEDAGGFSALIHSG	276	

RESULT 5  
B85518  
hypothetical protein Z0347 [imported] - Escherichia coli (strain O157:H7, substrain E  
C:/Species: Escherichia coli  
C:/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:/Accession: B85518  
R:/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
hiller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:/Accession number: A85480; MUID:21074935; PMID:11206551  
A:/Accession: B85518  
A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1-352 <570>  
A:/Cross-references: GR:AE005174; NID:g12513064; PIDD:MA654606.1; GSPDB:GN00145; UMG:  
A:/Experimental source: Strain O157:H7, substrain EDL933  
C:/Genetics:  
C:/Gene: Z0347

Query Match	46.4%	Score 45.5	DB 2	Length 352
Best Local Similarity	64.3%	Pred. No. 9.5		
Matches 9; Conservative	1;	Mismatches 3;	Indels 1;	Gaps 1.

RESULT 6  
F90667  
hypothetical protein ECs0310 [imported] - Escherichia coli (strain O157:H7, substrain R1)  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: F90667  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kuwakawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
Gasawara, N.; Yasunaga, T.; Kubota, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: F90667  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-378 <NAV>  
A:Cross-references: GB:BA000007; PIDN:BA033733.1; PID:q13359767; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain R1MD 0509952  
C:Genetics:  
A:Gene: ECs0310

Query Match 46.4%; Score 45.5; DB 2; Length 378;  
Best Local Similarity 64.3%; Pred. No. 10;  
Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

OY 4 WMDAESFGAIAHGC 17  
| | | | |  
Db 227 WKDAQS-GTFAHGC 239

RESULT 7  
AG0406  
conserved hypothetical protein YPO3349 [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C:Accession: AG0406  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AG0406  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-224 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC92579.1; PID:q15981276; GSPDB:GN00175  
C:Genetics:  
A:Gene: YPO3349

Query Match 45.9%; Score 45; DB 2; Length 224;  
Best Local Similarity 64.3%; Pred. No. 6.8;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 MDAESFGAIAHGC 18  
| | | | |  
Db 111 MDAVAFGMLKGL 124

RESULT 8  
AB3563  
conserved hypothetical protein PA0666 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: AB3563  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: AB2950; MUID:20437337  
A:Accession: AB3563

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-363 <STO>  
A:Cross-references: GB:AE004501; GB:AE004091; NID:g9946537; PIDN:AA04055.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0666  
C:Superfamily: hypothetical protein H10753

Query Match 45.9%; Score 45; DB 2; Length 363;  
Best Local Similarity 50.0%; Pred. No. 12;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 PAMMDAESFGAIAH 15  
| | | | |  
Db 322 PAMMDGMAFAMLAH 335

RESULT 9  
H87019  
probable glycosyl transferase, ML0886 [imported] - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: H87019  
R:Cole, S.T.; Eigemeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;  
R. Davies, R.M.; Devlin, K.; Dutfoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Hollo  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.;  
A:Title: Massive gene decay in the leprosy bacillus.  
A:Reference number: A86909; MUID:21128732; PMID:11234002  
A:Accession: H87019  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-384 <STO>  
A:Cross-references: GB:AL450380; NID:q13092958; PIDN:CAC31267.1; GSPDB:GN00147  
C:Genetics:  
A:Gene: ML0886

Query Match 45.9%; Score 45; DB 2; Length 384;  
Best Local Similarity 47.1%; Pred. No. 13;  
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 APAMMDAESFGAIAHGC 17  
| | | | |  
Db 41 APOMTGADTFDMAHGC 57

RESULT 10  
G72393  
hypothetical protein TM0306 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: G72393  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
Garret, M.M.; Stewart, A.M.; Cotton, M.D.; Pratl, M.S.; Phillips, C.A.; Richardson,  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A:Reference number: A72200; MUID:99287316  
A:Accession: G72393  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-449 <ARN>  
A:Cross-references: GB:AE001712; GB:AE000512; NID:g4980799; PIDN:AA035394.1; PID:g498  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0306  
C:Superfamily: alpha-L-fucosidase

Query Match 45.9%; Score 45; DB 2; Length 449;  
 Best Local Similarity 41.2%; Pred. No. 15;  
 Matches 7; Conservative 1; Mismatches 9; Indels 0; Gaps 0;  
 QY 2 PAMDAESFGAIAHGGL 18  
 Db 21 PKMFDAKFGIRIMGCI 37

RESULT 11  
 alpha-L-fucosidase (EC 3.2.1.51) precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 22-Jun-1999  
 C:Accession: S07074; S10235  
 R:Fisher, K.J.; Aronson Jr., N.N.  
 Biochem. J. 264, 695-701, 1989  
 A>Title: Isolation and sequence analysis of a cDNA encoding rat liver alpha-L-fucosidase  
 A:Reference number: S07074; MUID:90147522  
 A:Accession: S07074  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-376; 'R', 378-462 <FIS>  
 A:Cross-references: EMBL:X16145; NID:955650; PIDN:CAA34268.1; PID:955651  
 R:Aronson Jr., N.N.  
 Submitted to the EMBL Data Library, August 1989  
 A:Reference number: S10235  
 A:Accession: S10235  
 A:Molecule type: mRNA  
 A:Residues: 1-462 <ARO>  
 A:Cross-references: EMBL:X16145  
 C:Superfamily: alpha-L-fucosidase  
 C:Keywords: glycosidase; hydrolase

Query Match 45.9%; Score 45; DB 2; Length 462;  
 Best Local Similarity 41.2%; Pred. No. 16;  
 Matches 7; Conservative 1; Mismatches 9; Indels 0; Gaps 0;  
 QY 2 PAMDAESFGAIAHGGL 18  
 Db 46 PRMFDEAKFGLEFHVGV 62

RESULT 12  
 ABC transporter ATP binding protein ybab [imported] - Lactococcus lactis subsp. lactis  
 C:Species: Lactococcus lactis subsp. lactis  
 C>Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: D86637  
 R:Boletín, A.; Winkler, P.; Mauger, S.; Jallón, O.; Malarme, K.; Welschenbach, J.; Ehrlich  
 Genome Res. 11, 731-753, 2001  
 A>Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss  
 A:Reference number: A86625; MUID:21235186; PMID:11337471  
 A:Accession: D86637  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-517 <STO>  
 A:Cross-references: GB:AEO05176; PID:912722945; PIDN:AAK04198.1; GSPDB:GN00146  
 A:Experimental source: strain IL403  
 C:Genetics:  
 A:Gene: ybab  
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 44.9%; Score 44; DB 2; Length 517;  
 Best Local Similarity 52.9%; Pred. No. 26;  
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
 QY 2 PAMDAESFGAIAHGGL 18  
 Db 412 PSMWAELRGALAVGL 428

RESULT 13  
 hypothetical protein F22D6.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
 C:Accession: T21253  
 R:Wilkinson, J.  
 Submitted to the EMBL Data Library, April 1996  
 A:Reference number: Z19397  
 A:Accession: T21253  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-545 <NTL>  
 A:Cross-references: EMBL:Z71262; PIDN:CAA95808.1; GSPDB:GN00019; CESP:F22D6.3  
 A:Experimental source: clone F22D6  
 C:Genetics:  
 A:Gene: CESP:F22D6.3  
 A:Map position: 1  
 A:Introns: 28/3; 160/3; 318/2; 376/3; 430/3  
 C:Superfamily: lysine-tRNA ligase

Query Match 44.9%; Score 44; DB 2; Length 545;  
 Best Local Similarity 42.9%; Pred. No. 28;  
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 QY 4 WMDAESFGAIAHGGL 17  
 Db 499 YMDORKYGVPHGG 512

RESULT 14  
 hypothetical protein yhfC - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
 C:Accession: A69830  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.;  
 A.; Ehrlich, S.D.; Emerson, P.T.; Eutlian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Gallazzi, A.; Gal  
 lech, J.; Harwood, C.R.; Henselt, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M  
 Koetter, P.; Konigstein, G.; Krogg, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
 Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
 Rieger, M.; Rivoita, C.; Rochna, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
 akuchl, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Toso, V.; Uchida  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Yashima, A.  
 A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: A69830  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-258 <KUN>  
 A:Cross-references: GB:Z99109; GB:AL009126; NID:92633260; PIDN:CAB12658.1; PID:el1830  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yhfC

Query Match 44.4%; Score 43.5; DB 2; Length 258;  
 Best Local Similarity 60.0%; Pred. No. 14;  
 Matches 9; Conservative 1; Mismatches 4; Indels 1; Gaps 1;  
 QY 4 WMDAESFGAIAHGGL 18  
 Db 108 WADGLAFGA-GHGGL 121

RESULT 15

A11883

hypothetical protein alr0618 [imported] - Anabaena sp. (strain PCC 7120)

C;Species: Anabaena sp.

A;Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002

C;Accession: A11883

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuch

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: A11883

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-242 <KUR>

A;Cross-references: GB:BA000019; PIDN:BA072576.1; PID:q17129964; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr0618

Query Match

43.9%; Score 43; DB 2; Length 242;

Best Local Similarity 61.5%; Pred. No. 16;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 MDAESFGAIHGG 17

:1111111111

Db 218 LDAFKINAIAYGG 230

Search completed: July 8, 2002, 08:20:15  
Job time: 375 sec





US-08-932-376A-4  
; Sequence 4, Application US/08932376A  
; Patent No. 5869309  
; GENERAL INFORMATION:  
; APPLICANT: Politino, Michael  
; APPLICANT: Tonzi, Sean M.  
; APPLICANT: Usher, John J.  
; APPLICANT: Burnett K, William V.  
; APPLICANT: Romancik, Guna  
; TITLE OF INVENTION: CEPHALOSPORIN ESTERASE GENE FROM  
; TITLE OF INVENTION: RHODOSPORIDIUM TORULOIDES  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bristol-Myers Squibb Company  
; STREET: Rt. 206 & Provinceline Road  
; CITY: Princeton  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 08543-4000  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,376A  
; FILING DATE: 17-SEP-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Savitsky, Thomas R.  
; REGISTRATION NUMBER: 31,661  
; REFERENCE/DOCKET NUMBER: ON0144a  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (609) 252-4956  
; TELEFAX: (609) 252-4526  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 544 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-932-376A-4

Query Match 49.0%; Score 48; DB 2; Length 544;  
Best Local Similarity 52.9%; Pred. No. 4;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 PAMMDAESFGAIAHGCL 18  
| : ||| : |||  
DB 414 PTYWTAEAFGSSAHKGL 430

RESULT 3  
US-08-932-376A-2  
; Sequence 2, Application US/08932376A  
; Patent No. 5869309  
; GENERAL INFORMATION:  
; APPLICANT: Politino, Michael  
; APPLICANT: Tonzi, Sean M.  
; APPLICANT: Usher, John J.  
; APPLICANT: Burnett K, William V.  
; APPLICANT: Romancik, Guna  
; TITLE OF INVENTION: CEPHALOSPORIN ESTERASE GENE FROM  
; TITLE OF INVENTION: RHODOSPORIDIUM TORULOIDES  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bristol-Myers Squibb Company  
; STREET: Rt. 206 & Provinceline Road  
; CITY: Princeton  
; STATE: New Jersey  
; COUNTRY: USA

ZIP: 08543-4000  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,376A  
; FILING DATE: 17-SEP-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Savitsky, Thomas R.  
; REGISTRATION NUMBER: 31,661  
; REFERENCE/DOCKET NUMBER: ON0144a  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (609) 252-4956  
; TELEFAX: (609) 252-4526  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 572 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-932-376A-2

Query Match 49.0%; Score 48; DB 2; Length 572;  
Best Local Similarity 52.9%; Pred. No. 4.2;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 PAMMDAESFGAIAHGCL 18  
| : ||| : |||  
DB 442 PTYWTAEAFGSSAHKGL 458

US-09-199-637A-202  
; Sequence 202, Application US/09199637A  
; Patent No. 6355411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-Wah  
; APPLICANT: Cao, Hui  
; APPLICANT: Drenkard, Eliana  
; APPLICANT: Tsongalis, John  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A  
; CURRENT FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,517  
; PRIOR FILING DATE: 1997-11-25  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 202  
; LENGTH: 550  
; TYPE: PRN  
; ORGANISM: Pseudomonas aeruginosa  
; US-09-199-637A-202

Query Match 46.9%; Score 46; DB 4; Length 550;  
Best Local Similarity 53.3%; Pred. No. 8.6;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 APAMDAESFGAIAH 15  
| | | | : ||| : |||  
DB 34 AEAMSAALAYGALAH 48

RESULT 5  
US-09-199-637A-201  
; Sequence 201, Application US/09199637A  
; Patent No. 6355411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-Wah  
; APPLICANT: Cao, Hui  
; APPLICANT: Tsongalis, John  
; APPLICANT: Drenkard, Eliana  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A  
; PRIOR FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,517  
; PRIOR FILING DATE: 1997-11-25  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 201  
; LENGTH: 608  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-199-637A-201

Query Match 46.9%; Score 46; DB 4; Length 608;  
Best Local Similarity 53.3%; Pred. No. 9,6;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
OY 1 APAMDAESFGAIAH 15  
| | | : : : | | |  
Db 92 AEAMSAAGALAH 106  
RESULT 6  
US-09-199-637A-200  
; Sequence 200, Application US/09199637A  
; Patent No. 6355411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-Wah  
; APPLICANT: Cao, Hui  
; APPLICANT: Drenkard, Eliana  
; APPLICANT: Tsongalis, John  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A  
; PRIOR FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,517  
; PRIOR FILING DATE: 1997-11-25  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 200  
; LENGTH: 639  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-199-637A-200

Query Match 46.9%; Score 46; DB 4; Length 639;  
Best Local Similarity 53.3%; Pred. No. 10;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
OY 1 APAMDAESFGAIAH 15  
| | | : : : | | |

Db 123 AEAMSAAGALAH 137  
RESULT 7  
US-09-199-637A-103  
; Sequence 103, Application US/09199637A  
; Patent No. 6355411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-Wah  
; APPLICANT: Cao, Hui  
; APPLICANT: Drenkard, Eliana  
; APPLICANT: Tsongalis, John  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A  
; PRIOR FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,517  
; PRIOR FILING DATE: 1997-11-25  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 103  
; LENGTH: 641  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-199-637A-103

Query Match 46.9%; Score 46; DB 4; Length 641;  
Best Local Similarity 53.3%; Pred. No. 10;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
OY 1 APAMDAESFGAIAH 15  
| | | : : : | | |  
Db 129 AEAMSAAGALAH 143

RESULT 8  
US-09-199-637A-199  
; Sequence 199, Application US/09199637A  
; Patent No. 6355411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-Wah  
; APPLICANT: Cao, Hui  
; APPLICANT: Drenkard, Eliana  
; APPLICANT: Tsongalis, John  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A  
; PRIOR FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,517  
; PRIOR FILING DATE: 1997-11-25  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 199  
; LENGTH: 643  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-199-637A-199

Query Match 46.9%; Score 46; DB 4; Length 643;  
Best Local Similarity 53.3%; Pred. No. 10;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 APAMMDAESFGAIAH 15  
||| | :|||  
Db 127 AEWSAAMAYGALAH 141

## RESULT 9

US-09-199-637A-253  
; Sequence 253, Application US/09199637A  
; Patent No. 6355411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rathee, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Cao, Hui  
; APPLICANT: Drenkard, Eliana  
; APPLICANT: Tsongalis, John  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A  
; PRIOR FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,517  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 253  
; LENGTH: 645  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-199-637A-253

Query Match 46.9%; Score 46; DB 4; Length 645;  
Best Local Similarity 53.3%; Pred. No. 10;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 APAMMDAESFGAIAH 15  
||| | :|||  
Db 129 AEWSAAMAYGALAH 143

## RESULT 10

US-08-868-288A-3  
; Sequence 3, Application US/08868288A  
; Patent No. 5922567  
; GENERAL INFORMATION:  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Lal, Preeti  
; APPLICANT: Bandman, Olga  
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/868,288A  
; FILING DATE: June 3, 1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0309 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0553  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 330 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: HNT2RAT01  
CLONE: 260873  
US-08-868-288A-3

Query Match 43.9%; Score 43; DB 2; Length 330;  
Best Local Similarity 70.0%; Pred. No. 15;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 9 SFGAIAHGCL 18  
||| : |||  
Db 163 SFGSLGHGCL 172

## RESULT 11

US-09-235-373-3  
; Sequence 3, Application US/09235373  
; Patent No. 6001598  
; GENERAL INFORMATION:  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Lal, Preeti  
; APPLICANT: Bandman, Olga  
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/235,373  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/868,288  
; FILING DATE: June 3, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0309 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0553  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 330 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: HNT2RAT01  
CLONE: 260873  
US-09-235-373-3

Query Match 43.9%: Score 43; DB 3; Length 330;  
Best Local Similarity 70.0%; Pred. No. 15;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 9 SFGAIAHGGL 18  
|||||  
Db 163 SFGSLGHGGL 172

RESULT 12

US-09-388-993-3  
; Sequence 3, Application US/09388993  
; Patent No. 6043222

GENERAL INFORMATION:

APPLICANT: Au-Yang, Janice

APPLICANT: Lal, Preeti

APPLICANT: Bandman, Olga

TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/388,993

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/868,288

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0309 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 330 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: HMT2RAT01

CLONE: 260873

US-09-388-993-3

QY 9 SFGAIAHGGL 18  
|||||  
Db 163 SFGSLGHGGL 172

Query Match 43.9%: Score 43; DB 3; Length 330;  
Best Local Similarity 70.0%; Pred. No. 15;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

RESULT 13

US-08-598-873-58

; Sequence 58, Application US/08598873

; Patent No. 5928884

GENERAL INFORMATION:

APPLICANT: Croce, Carlo M.

APPLICANT: Huebner, Kay

TITLE OF INVENTION: FHIT PROTEINS AND NUCLEIC ACIDS AND  
METHODS BASED THEREON  
NUMBER OF SEQUENCES: 77  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/598,873

FILING DATE: 09-FEB-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Friedel, Thomas E.

REGISTRATION NUMBER: 29,258

REFERENCE/DOCKET NUMBER: 8666-004

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:

LENGTH: 122 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-598-873-58

QY 1 APAMDAESF 10  
|||  
Db 75 SPWODAOSEF 84

Query Match 40.8%: Score 40; DB 2; Length 122;  
Best Local Similarity 60.0%; Pred. No. 16;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RESULT 14

US-08-605-430-58

; Sequence 58, Application US/08605430

; Patent No. 6242212

GENERAL INFORMATION:

APPLICANT: Croce, Carlo M.

APPLICANT: Huebner, Kay

TITLE OF INVENTION: FHIT PROTEINS AND NUCLEIC ACIDS AND

METHODS BASED THEREON

NUMBER OF SEQUENCES: 86

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/605,430

FILING DATE: 22-FEB-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Friedel, Thomas E.

REGISTRATION NUMBER: 29,258  
REFERENCE/DOCKET NUMBER: 8666-005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 780-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 122 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-605-430-58

Query Match 40.8%; Score 40; DB 4; Length 122;  
Best Local Similarity 60.0%; Pred. No. 16;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 APAMDAESF 10  
: | | | | |  
Db 75 SPQWQDAQSF 84

RESULT 15  
US-08-844-154-2  
Sequence 2, Application US/08844154  
Patent No. 5827708  
GENERAL INFORMATION:  
APPLICANT: Lawlor, Elizabeth  
TITLE OF INVENTION: No. 5827708e1 Compounds  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithkline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/844,154  
FILING DATE: 18-APR-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9607993.4  
FILING DATE: 18-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimm, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P31457-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 434 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-844-154-2

Query Match 40.8%; Score 40; DB 2; Length 434;  
Best Local Similarity 35.7%; Pred. No. 64;  
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 4 WMDAESGAIHGC 17  
: | | | | |  
Db 401 YLDLRKRYGTVPHEG 414

Search completed: July 8, 2002, 08:19:40  
Job time: 585 sec

Mon Jul 8 09:33:13 2002

us-09-582-296-1.ra1

1

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:09:35 ; Search time 51.87 Seconds  
(without alignments)  
38.545 Million cell updates/sec

Title: us-09-582-296-1

Perfect score: 98  
Sequence: 1 APAMMDAEEFGAIAHGCL 18

Scoring table: BLOSUM62  
Gapop 10.0 , Capext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq.032802.\*

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	98	100.0	18 20 AAY27010	N-terminal fragmen
2	69	70.4	20 20 AAY27011	N-terminal fragmen
3	50	51.0	15 AAR56285	Sequence of human
4	49	50.0	129 22 AAB85247	Human secreted pro
5	49	50.0	184 21 AAB14315	Human secreted pro
6	49	50.0	22 AAB85237	Human secreted pro
7	49	50.0	19 AAW75057	Human secreted pro
8	49	50.0	20 AAY13578	Amino acid sequenc
9	49	50.0	21 AAY94860	Human protein cion
10	49	50.0	22 AAU29031	Human PRO polypept
11	49	50.0	22 AAW78342	Human protein SEO

12	49	50.0	467 22 AAB88401	Human membrane or
13	49	50.0	467 22 AAB80246	Human PRO260 prote
14	49	50.0	496 22 AAM79326	Human protein SEQ
15	48	49.0	572 19 AAM54159	Rhodospiridium tor
16	48	49.0	572 22 AAB82871	Rhodospiridium tor
17	46	46.9	296 21 AAY91613	Human secreted pro
18	46	46.9	334 22 AAU61259	Propionibacterium
19	46	46.9	383 22 ABB11745	Human netrin 4 hom
20	46	46.9	549 20 AAY29270	Protein encoded by
21	46	46.9	607 20 AAY29269	Protein encoded by
22	46	46.9	628 20 AAG66449	Human beta-netrin.
23	46	46.9	639 20 AAY29268	Protein encoded by
24	46	46.9	642 20 AAY29266	Virulence factor p
25	46	46.9	643 20 AAY29267	Protein encoded by
26	46	46.9	645 20 AAY29172	Amino acid sequenc
27	46	46.9	122 22 AAY29615	Novel human secret
28	44	44.9	608 22 AAG91551	C glutamicum prote
29	43	43.9	317 20 AAY74126	Human prostate tum
30	43	43.9	330 20 AAM94066	Human DnaD-like pr
31	43	43.9	468 22 AAM38945	Human novel protei
32	43	43.9	468 22 AAU14130	Human novel protei
33	43	43.9	607 22 AAG63535	Propionibacterium
34	42.5	43.4	280 22 AAG90751	C glutamicum prote
35	42.5	43.4	1222 22 ABB67050	Drosophila melanog
36	42.5	43.4	2506 22 ABB67191	Novel human diagno
37	42.5	43.4	2506 22 ABB67191	Novel human diagno
38	42	42.9	74 22 AAU42894	Novel human diagno
39	42	42.9	591 22 AAU76222	Pseudomonas aerugi
40	42	42.9	700 22 AAE02420	Frog vasa protein.
41	42	42.9	899 22 ABB65488	Drosophila melanog
42	42	42.9	899 22 ABB65489	Drosophila melanog
43	41	41.8	93 21 AAB40640	Human ORFX ORF404
44	41	41.8	111 22 ABB22358	Novel human diagno
45	41	41.8	121 22 AA003564	Human polypeptide

#### ALIGNMENTS

RESULT 1	
AAV27010	standard; peptide: 18 AA.
ID	AAV27010;
AC	24-SEP-1999 (first entry)
XX	
DT	N-terminal fragment of anti-freeze protein from Lichen.
XX	
DE	Antifreeze protein; AFP; Lichen; frozen food product; pasteurisation;
XX	
KW	sterilisation; freezing; frozen confectionery product; N-terminal.
XX	
OS	Umbilicaria antarctica.
XX	
PN	WO9937673-A2.
XX	
PD	29-JUL-1999.
XX	
PF	23-DEC-1998; 98WO-EP08554.
XX	
PR	22-JAN-1998; 98GB-0001420.
XX	
PA	(UNTL) UNILEVER NV.
PA	(UNTL) UNILEVER PLC.
PI	Byass LJ, Sidebottom CM, Smallwood MF;
XX	
DR	WPI, 1999-444595/37.
PT	New isolated antifreeze protein obtained from Lichen, used for the
PT	preparation of food products, particularly frozen confectionery
PT	products
XX	

PS Claim 1; Page 17; 20pp; English.

XX The invention describes a novel antifreeze protein (AFP) obtained from  
CC Lichen that comprises an apparent mol. wt. of 20-28kD and has an  
CC N-terminal amino acid sequence which shows at least 80% overlap with  
CC the present sequence; and modified versions and isoforms of this  
CC protein. The AFP can conveniently be used in food products, preferably  
CC in food products which are frozen or intended to be frozen. Especially  
CC pasteurisation or sterilisation prior to freezing and in frozen  
CC confectionery products. Using the AFP ingredient, mixes can be frozen  
CC under quiescent conditions, e.g. in a shop or home freezer without the  
CC formation of unacceptable ice crystal shapes and hence with a texture  
CC different to products normally obtained via quiescent freezing. The  
CC present sequence represents a N-terminal fragment of an AFP.

SQ Sequence 18 AA;

Query Match 100.0%; Score 98; DB 20; Length 18;  
Best Local Similarity 100.0%; Pred. No. 1.4e-09;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 APAMDAESFGATAHGGL 18  
Db 1 apawmdaesfgatahgg1 18

# RESULT 2

AAV27011  
ID AAV27011 standard; peptide; 20 AA.

AC AAV27011;

DT 24-SEP-1999 (first entry)

XX N-terminal fragment of anti-freeze protein from Lichen.

KW Antifreeze protein; AFP; Lichen; frozen food product; pasteurisation;  
KM sterilisation; freezing; frozen confectionery product; N-terminal.

OS Umbilicaria antarctica.

PN WO937673-A2.

PD 29-JUL-1999.

PF 23-DEC-1998; 98WO-EP08554.

PR 22-JAN-1998; 98GB-0001420.

PA (UNIL ) UNILEVER NV.  
PA (UNIL ) UNILEVER PLC.

PI Byass LJ, Sidebottom CM, Smallwood MF;

DR WPI; 1999-444595/37.

PT New isolated antifreeze protein obtained from Lichen, used for the  
PT preparation of food products, particularly frozen confectionery  
PT products

PS Claim 2; Page 17; 20pp; English.

XX The invention describes a novel antifreeze protein (AFP) obtained from  
CC Lichen that comprises an apparent mol. wt. of 20-28kD and has an  
CC N-terminal amino acid sequence which shows at least 80% overlap with  
CC sequence shown in AAV27010; and modified versions and isoforms of  
CC this protein. The AFP can conveniently be used in food products,  
CC preferably in food products which are frozen or intended to be frozen.  
CC Especially preferred is the use of AFPs in products which are heated e.g.  
CC by pasteurisation or sterilisation prior to freezing and in frozen  
CC confectionery products. Using the AFP ingredient, mixes can be frozen

CC under quiescent conditions, e.g. in a shop or home freezer without the  
CC formation of unacceptable ice crystal shapes and hence with a texture  
CC different to products normally obtained via quiescent freezing. The  
CC present sequence represents a N-terminal fragment of an 24 kDa AFP.

SQ Sequence 20 AA;

Query Match 70.4%; Score 69; DB 20; Length 20;  
Best Local Similarity 80.0%; Pred. No. 0.0001;  
Matches 16; Conservative 1; Mismatches 1; Indels 2; Gaps 1;

OY 1 APAM--DAESFGATAHGGL 18  
Db 1 apavmmdaesfgatahgg1 20

# RESULT 3

AAR56285  
ID AAR56285 standard; Protein; 461 AA.

AC AAR56285;

DT 03-MAR-1995 (first entry)

XX Sequence of human alpha-L-fucosidase including signal peptide.

XX Anti-adhesion enzyme; alpha-L-fucosidase.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..22

FT Protein /label= signal

FT /label= mature

FT WO9414472-A.

PD 07-JUL-1994.

PF 21-DEC-1993; 93WO-US12464.

PR 22-DEC-1992; 92US-0994650.

PA (REGC ) UNIV CALIFORNIA.

PI Carson DA, Wicks I;

DR WPI; 1994-234351/28.

DR N-PSDB; AA065537.

PT Inhibiting adhesion of cells to selectin-contg. receptors - by  
PT treatment with enzyme degrading carbohydrate residues in selectin  
PT adhesion ligands, partic. for treating inflammation

PS Disclosure; Page 27-30; 42pp; English.

XX The AA sequence in R56285 is derived from sequence information obtd.  
CC through isolation of several human alpha-L-fucosidase cDNA clones  
CC reported by Occhiodoro et al. at Biochem. Biophys. Res. Comm.,  
CC 164:439-445 (1989). The full-length nucleotide sequence for alpha-L-  
CC fucosidase is available from the GENBANK molecular sequence database  
CC under Accession No. M28099. The DNA sequence in Q6537 is described  
CC as genomic.

SQ Sequence 461 AA;

Query Match 51.0%; Score 50; DB 15; Length 461;  
Best Local Similarity 47.1%; Pred. No. 5.3;  
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;



OS	Homo sapiens.
XX	
PN	WO200144267-A2.
XX	
PD	21-JUN-2001.
XX	
PF	18-DEC-2000; 2000WO-US34433.
XX	
PR	16-DEC-1999; 99US-0465512.
PR	17-DEC-1999; 99WO-US29854.
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shinkets RA;
XX	
DR	WPI: 2001-408469/43.
DR	N-PSTDB; AAM22700.
XX	
PT	New secreted human protein, SECX is useful for the treatment of
PT	pathological disorders i.e. cancer -
XX	
PS	
XX	
CC	Claim 17; Fig 16A; 162pp; English.
CC	The invention provides novel secreted human proteins (SECX) and
CC	nucleotide sequences encoding the SECX polypeptides. The SECX
CC	polypeptides can be expressed by standard recombinant methodology. SECX
CC	polypeptides are useful for the treatment, prevention or delay of onset
CC	of a pathological condition associated with aberrant SECX expression or
CC	activity, i.e. cancer. The present sequence represents the amino acid
CC	of a human SECX polypeptide (see DE line for clone identification).
XX	
SQ	Sequence 184 AA;
Query Match	50.0%; Score 49; DB 22; Length 184;
Best Local Similarity	47.1%; Pred No. 2.7;
Matches 8; Conservativity 1; Mismatches 8; Indels 0; Gaps 0	
OY	2 PAMMDAESFGAIHAGCL 18             47 pawfdqakfglfihgv 63
Dd	
RESULT 7	
AAM75057	
ID	AAM75057 standard; Protein; 467 AA.
XX	
AC	AAM75057;
XX	
DT	28-JAN-1999 (first entry)
XX	
DE	Human secreted protein encoded by gene 1 clone HGCMD20.
XX	
RW	Human; secreted protein; fusion protein; gene therapy; protein therapy;
RW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
RW	developmental abnormality; foetal deficiency; blood; allergy; renal;
RW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
RW	inflammation; ischaemic shock; Alzheimer's disease; testostosis; AIDS;
RW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
RW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
RW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 97 /label= unknown
FT	Misc-difference 119 /label= unknown
FT	Misc-difference 240 /label= unknown
XX	
FN	WO9839446-A2

[illegible]



PA	(GETH ) GENENTECH INC.
XX	
PI	Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
XI	
DR	WPI; 1999-229537/19.
XX	N-PSDB; AAX52249.
PT	New isolated human genes and polypeptides used in, e.g. treatment of
XX	gastrointestinal ulceration
PS	Claim 12; Fig 72; 320pp; English.
CC	AAY13344-403 represent secreted and transmembrane human proteins.
CC	The CDNA sequences are obtained from cDNA libraries, prepared from
CC	fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
CC	The encoded polypeptides have specific uses based on their homology to
CC	known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
CC	associated with the preservation and maintenance of gastrointestinal
CC	mucosa and the repair of acute and chronic mucosal lesions
CC	(e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
CC	ulceration and congenital microvillus atrophy), skin diseases associated
CC	with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
CC	cancers such as lung squamous cell carcinoma of the vulva and gliomas),
CC	potent effects on cell growth and development, diseases related to
CC	growth or survival of nerve cells including Parkinson's disease,
CC	Alzheimer's disease, AIDS, neuropathies or cancer. PRO265 can be used as
CC	as a target for anti-tumor drugs. PRO533 may be used in the treatment
CC	of Usher Syndrome or Atrophia areata; PRO269 can be used as an
CC	anti-thrombotic agent; PRO287 polypeptides and portions may have
CC	therapeutic applications in wound healing and tissue repair; PRO317 can
CC	be used for treating problems of the kidney, uterus, endometrium, blood
CC	vessels, or related tissue, e.g. in the heart of genital tract.
SQ	
XX	Sequence 467 AA;
OY	2 PAMMDAEEFGAIHAGCL 18            : Db 47 pawfdqakfgilhwgv 63
RESULT 9	
AAY94860	
ID AAY94860 standard; Protein; 467 AA.	
XX AC	
XX AAY94860;	
XX 12-JUN-2000 (first entry)	
XX Human protein clone HP02575.	
XX	
XX Homo sapiens.	
OS	
XX WO200005367-A2.	
PN	
XX 03-FEB-2000.	

[illegible]

AC AAU29031:  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Human PRO polypeptide sequence #8.  
 XX  
 KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;  
 KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;  
 KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;  
 KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200168848-A2.  
 XX  
 PD 20-SEP-2001.  
 XX  
 PF 28-FEB-2001; 2001WO-US06520.  
 XX  
 PR 01-MAR-2000; 2000WO-US05601.  
 PR 02-MAR-2000; 2000WO-US05641.  
 PR 03-MAR-2000; 2000US-187202P.  
 PR 06-MAR-2000; 2000US-186968P.  
 PR 14-MAR-2000; 2000US-189320P.  
 PR 14-MAR-2000; 2000US-189328P.  
 PR 15-MAR-2000; 2000WO-US06884.  
 PR 21-MAR-2000; 2000US-190828P.  
 PR 21-MAR-2000; 2000US-191007P.  
 PR 21-MAR-2000; 2000US-191048P.  
 PR 21-MAR-2000; 2000US-191314P.  
 PR 28-MAR-2000; 2000US-192655P.  
 PR 29-MAR-2000; 2000US-193032P.  
 PR 29-MAR-2000; 2000US-193053P.  
 PR 30-MAR-2000; 2000WO-US08439.  
 PR 04-APR-2000; 2000US-194449P.  
 PR 11-APR-2000; 2000US-195975P.  
 PR 11-APR-2000; 2000US-196000P.  
 PR 11-APR-2000; 2000US-196187P.  
 PR 11-APR-2000; 2000US-196690P.  
 PR 11-APR-2000; 2000US-196820P.  
 PR 18-APR-2000; 2000US-198121P.  
 PR 18-APR-2000; 2000US-198585P.  
 PR 25-APR-2000; 2000US-199397P.  
 PR 25-APR-2000; 2000US-199550P.  
 PR 25-APR-2000; 2000US-199654P.  
 PR 03-MAY-2000; 2000US-201516P.  
 PR 17-MAY-2000; 2000WO-US13705.  
 PR 22-MAY-2000; 2000WO-US14042.  
 PR 30-MAY-2000; 2000WO-US14941.  
 PR 02-JUN-2000; 2000WO-US15264.  
 PR 05-JUN-2000; 2000US-209832P.  
 PR 28-JUL-2000; 2000WO-US20710.  
 PR 22-AUG-2000; 2000US-0644848.  
 PR 24-AUG-2000; 2000WO-US23328.  
 PR 08-NOV-2000; 2000WO-US30952.  
 PR 01-DEC-2000; 2000WO-US32658.  
 PR 20-DEC-2000; 2000WO-US34956.  
 XX  
 PA (GENTH) GENENTECH INC.  
 XX  
 PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;  
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;  
 XX  
 DR WPI; 2001-602746/68.  
 DR N-PSDB; AAS45932.  
 XX  
 PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the  
 PT presence of tumours, such as prostate and breast tumours, in mammals and  
 PT to screen for modulators of the compounds -  
 XX  
 PS Claim 11; Fig 16; 774pp; English.  
 XX

CC Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.  
 CC The PRO polypeptides and their associated nucleic acids can be used to  
 CC detect the presence of a tumour in a mammal by comparing the level of  
 CC expression of a PRO polypeptide in a test sample of cells from the animal  
 CC and a control sample of normal cells, whereby a higher level of  
 CC expression in the test sample indicates the presence of a tumour in the  
 CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats  
 CC and rabbits but are preferably human. The polypeptides can be used to  
 CC stimulate tumour necrosis factor (TNF) alpha release from human blood,  
 CC when contacted with it. A specific polypeptide can be used to stimulate  
 CC the proliferation or differentiation of chondrocyte cells. The PRO  
 CC proteins can be used to determine the presence of tumours and also  
 CC susceptibility to tumour development, particularly adrenal, lung, colon,  
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian  
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids  
 CC can be used for genetic analysis of individuals with genetic disorders.  
 XX  
 SQ Sequence 467 AA:  
 XX  
 QY 2 PAMDAESFGAIAHGGL 18  
 Db 47 pawfdqakfgifhwgv 63  
 XX  
 RESULT 11  
 AAU78342  
 ID AAU78342 standard; Protein; 467 AA.  
 XX  
 AC AAU78342;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human protein SEQ ID NO 1004.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US04098.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Gao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX  
 DR WPI; 2001-476283/51.  
 DR N-PSDB; AAK51475.  
 XX  
 PT Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX

PS Claim 20; Page 3230-3231; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAW76323-AAW80302) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAW80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.

XX  
 SQ Sequence 467 AA;

Query Match 50.0%; Score 49; DB 22; Length 467;  
 Best Local Similarity 47.1%; Pred. No. 7.9;  
 Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 2 PAMMDAESFGAIAHGGL 18  
 DB 47 pawfdqakigfihwgv 63

## RESULT 12

AAW88401  
 ID AAB88401 standard; Protein: 467 AA.

AC AAB88401;

DT 23-MAY-2001 (first entry)

XX Human membrane or secretory protein clone PSEC0151.

KW Human: secretory protein; membrane protein; vaccine; gene therapy;

XX rheumatoid arthritis; diabetes.

OS Homo sapiens.

PN EPI067182-A2.

PD 10-JAN-2001.

XX

PF 07-JUL-2000; 2000EP-0114090.

PR 08-JUL-1999; 99JP-0194179.

PR 11-JAN-2000; 2000JP-0118775.

PR 02-MAY-2000; 2000JP-0183766.

XX

PA (HELI-) HELIX RES INST.

XX

PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;

DR WPI; 2001-093989/11.

DR N-PSDB; AAF93828.

XX

PT Nucleic acids encoding secretory proteins/membrane proteins, useful in

PT gene therapy or as candidate target molecules in drug development -

XX

PS Claim 1; SEQ ID 170; 609pp + CD ROM; English.

XX

CC This invention relates to nucleic acid sequences AAF93744 - AAF93916

CC which encode human secretory or membrane proteins represented by

CC AAB88317 - AAB88419. Included in the invention are primers

CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the

CC cDNA sequences of the invention. The invention also includes methods for

CC the production of antibodies directed against the proteins, and cDNA

CC sequences, which can be used in vaccines. The polynucleotide sequences

CC can be used in gene therapy. The polynucleotide sequences and the  
 CC proteins they encode may be used in the prevention, treatment and  
 CC diagnosis of diseases associated with inappropriate secretory  
 CC protein/membrane protein expression. The nucleic acids and complementary  
 CC sequences may also be used as DNA probes in diagnostic assays  
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantify the  
 CC presence of similar nucleic acid sequences in samples. They may also be  
 CC used to study the expression and function of secretory proteins/membrane  
 CC polypeptides and their role in metabolism. The polypeptides may be used  
 CC as antigens in the production of antibodies against them and in assays to  
 CC identify modulators (agonists and antagonists) of expression and  
 CC activity. The antibodies and antagonists may also be used as therapeutic  
 CC agents to down regulate expression and activity. The antibodies may also  
 CC be used as diagnostic agents for detecting the presence of the  
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay  
 CC (ELISA)). Examples of diseases which may be treated include rheumatoid  
 CC arthritis and diabetes.

XX  
 SQ Sequence 467 AA;

Query Match 50.0%; Score 49; DB 22; Length 467;  
 Best Local Similarity 47.1%; Pred. No. 7.9;  
 Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 2 PAMMDAESFGAIAHGGL 18  
 DB 47 pawfdqakigfihwgv 63

## RESULT 13

AAW80246  
 ID AAB80246 standard; Protein: 467 AA.

AC AAB80246;

DT 24-APR-2001 (first entry)

XX Human PRO260 protein.

KW Human: PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;

XX antiparkinsonian nootropic; neuroprotective; vulnary; cardiac;

XX antiangiogenic; vasotropic; antidiabetic; antirheumatic; cancer;

XX antiarthritic; antifertility; antidiabetic; antiviral; diabetes;

XX ophthalmological; gene therapy; skin disease; gastrointestinal disorder;

XX ischaemia; inflammation.

OS Homo sapiens.

PN WO200104311-A1.

PD 18-JAN-2001.

XX

PF 22-FEB-2000; 2000WO-US04414.

PR 07-JUL-1999; 99US-0143048.

PR 26-JUL-1999; 99US-0145698.

PR 08-SEP-1999; 99US-0146222.

PR 13-SEP-1999; 99WO-US20944.

PR 15-SEP-1999; 99WO-US21090.

PR 15-SEP-1999; 99WO-US21547.

PR 05-OCT-1999; 99WO-US23089.

PR 29-NOV-1999; 99WO-US28214.

PR 30-NOV-1999; 99WO-US28313.

PR 16-DEC-1999; 99WO-US30095.

PR 20-DEC-1999; 99WO-US30911.

PR 20-DEC-1999; 99WO-US30999.

PR 05-JAN-2000; 99WO-US00219.

XX

PA (GENT ) GENENTECH INC.

XX

PI Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;



DR WPI: 1998-230321/20.  
DR N-PSDB: AAV21460-61.

XX Isolated cephalosporin esterase gene - obtained from Rhodospiridium  
PT toruloides, used to obtain products which can be used for hydrolysis  
PT of cephalosporins

XX PS Claim 13: Fig 5: 38pp; English.

CC This protein comprises Rhodospiridium toruloides ATCC 10657  
CC cephalosporium esterase, an enzyme that catalyses the hydrolysis of  
CC the 3' acetyl groups of cephalosporins. The enzyme, especially the  
CC mature polypeptide, can be purified from cultures of R. toruloides  
CC or expressed in claimed host cells (preferably Escherichia coli, R.  
CC toruloides, Cephalosporium acremonium or Penicillium chrysogenum)  
CC transformed with a vector comprising cephalosporin esterase cDNA or  
CC genomic DNA (see AAV21460-61). The isolated polypeptides, even if  
CC catalytically inactive, can also be used for the production of  
CC antibodies for use in detection and purification.

XX SQ Sequence 572 AA;

Query Match 49.0%; Score 48; DB 19; Length 572;  
Best Local Similarity 52.9%; Pred. No. 15;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 PAMDAEFGAIAHGGL 18  
I : ||:||: || ||  
Db 442 plywtaefgssahkgl 458

Search completed: July 8, 2002, 08:19:12  
Job time: 577 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:19:40 ; Search time 21.46 Seconds  
(without alignments)  
22.764 Million cell updates/sec

Title: US-09-582-296-2

Perfect score: 101

Sequence: 1 APAYVMGDAESFGAIHGGI 20

Scoring table: BLOSUM62  
Gapop 10.0 , Capext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	44.6	617	1	US-08-279-700-21
2	43	42.6	330	2	US-08-868-288A-3
3	43	42.6	330	3	US-09-235-373-3
4	43	42.6	330	3	US-09-388-993-3
5	43	42.6	617	4	US-09-230-944-11
6	43	42.6	617	1	US-08-279-700-2
7	43	42.6	617	1	US-08-279-700-4
8	43	42.6	617	1	US-08-279-700-6
9	43	42.6	617	1	US-08-279-700-8
10	43	42.6	617	1	US-08-279-700-10
11	43	42.6	617	1	US-08-279-700-12
12	43	42.6	617	1	US-08-279-700-14
13	43	42.6	617	1	US-08-348-891A-6
14	43	42.6	617	2	US-08-905-817-6
15	43	42.6	617	2	US-09-230-944-10
16	41	40.6	544	2	US-08-932-376A-4
17	41	40.6	572	2	US-08-932-376A-2
18	40	39.6	247	1	US-08-446-083-5
19	39.5	39.1	388	1	US-08-290-448A-80
20	39.5	39.1	388	1	US-08-290-448A-80
21	39.5	39.1	388	1	US-08-175-069A-80
22	39.5	39.1	388	4	US-08-461-939B-80
23	39.5	39.1	388	4	US-08-464-000-80
24	39	38.6	426	6	5221737-2
25	38	37.6	151	2	US-08-485-455D-51
26	38	37.6	151	2	US-08-482-130C-51
27	38	37.6	151	2	US-08-484-211C-51

28	38	37.6	151	3	US-08-906-769-51	Sequence 51, Appl
29	38	37.6	151	3	US-08-906-616-51	Sequence 51, Appl
30	38	37.6	151	4	US-08-817-795-51	Sequence 51, Appl
31	38	37.6	151	4	US-08-485-443B-51	Sequence 51, Appl
32	38	37.6	151	4	US-08-639-075A-51	Sequence 51, Appl
33	38	37.6	151	4	US-09-012-431-51	Sequence 51, Appl
34	38	37.6	151	4	US-09-012-692-51	Sequence 51, Appl
35	38	37.6	151	4	US-08-906-613-51	Sequence 51, Appl
36	38	37.6	151	5	PCT-US95-14442A-51	Sequence 51, Appl
37	38	37.6	344	1	US-08-180-209B-58	Sequence 58, Appl
38	38	37.6	344	4	US-08-474-853-58	Sequence 58, Appl
39	38	37.6	344	5	PCT-US94-02629-58	Sequence 58, Appl
40	38	37.6	496	3	US-08-906-769-113	Sequence 113, App
41	38	37.6	496	3	US-08-906-616-113	Sequence 113, App
42	38	37.6	496	4	US-08-817-795-113	Sequence 113, App
43	38	37.6	496	4	US-08-639-075A-113	Sequence 113, App
44	38	37.6	496	4	US-09-012-431-113	Sequence 113, App
45	38	37.6	496	4	US-09-012-692-113	Sequence 113, App

#### ALIGNMENTS

RESULT 1  
US-08-279-700-21  
Sequence 21, Appl  
Patent No. 5578448  
GENERAL INFORMATION:  
APPLICANT: KOTA, Jennifer S.  
TITLE OF INVENTION: WILD-TYPE MEASLES VIRUS GLYCOPROTEINS:  
TITLE OF INVENTION: VACCINE AND DETECTION METHOD THEREFOR  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/279,700  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/866,033A  
FILING DATE: 19920408  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 40399/140 NIHD  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 617 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
SPRAIN: consensus HA polypeptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 4  
OTHER INFORMATION: /note= "Xaa denotes Gln or His"

FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 19  
OTHER INFORMATION: /note= "Xaa denotes Lys or Arg"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 176  
OTHER INFORMATION: /note= "Xaa denotes Thr, Val or  
OTHER INFORMATION: Ala"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 235  
OTHER INFORMATION: /note= "Xaa denotes Glu or Gly"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 295  
OTHER INFORMATION: /note= "Xaa denotes Lys or Arg"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 303  
OTHER INFORMATION: /note= "Xaa denotes Glu or Gly"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 305  
OTHER INFORMATION: /note= "Xaa represents Ser or Phe"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 306  
OTHER INFORMATION: /note= "Xaa denotes Ile or Val"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 308  
OTHER INFORMATION: /note= "Xaa denotes Ile or Val"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 320  
OTHER INFORMATION: /note= "Xaa denotes Gln or Arg"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 339  
OTHER INFORMATION: /note= "Xaa denotes Leu or Phe"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 348  
OTHER INFORMATION: /note= "Xaa denotes Arg or Lys"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 367  
OTHER INFORMATION: /note= "Xaa denotes Val or Ile"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 389  
OTHER INFORMATION: /note= "Xaa denotes Lys or Arg"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 390  
OTHER INFORMATION: /note= "Xaa denotes Ile or Asn"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 446  
OTHER INFORMATION: /note= "Xaa denotes Ser or Thr"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 451  
OTHER INFORMATION: /note= "Xaa denotes Val or Glu"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 485  
OTHER INFORMATION: /note= "Xaa denotes Val or Ile"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 501  
OTHER INFORMATION: /note= "Xaa denotes Pro or Ser"

FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 544  
OTHER INFORMATION: /note= "Xaa denotes Ser or Asn"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 546  
OTHER INFORMATION: /note= "Xaa denotes Ser or Gly"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 559  
OTHER INFORMATION: /note= "Xaa denotes Ile or Val"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 560  
OTHER INFORMATION: /note= "Xaa denotes Lys or Arg"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 562  
OTHER INFORMATION: /note= "Xaa denotes Val, Ile or  
OTHER INFORMATION: Phe"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 593  
OTHER INFORMATION: /note= "Xaa denotes His or Tyr"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 616  
OTHER INFORMATION: /note= "Xaa denotes Arg or Ser"  
US-08-279-700-21

Query Match 44.6%; Score 45; DB 1; Length 617;  
Best Local Similarity 50.0%; Pred. No. 9.3;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
QY 5 VMGDAESFGAIAHGCL 20  
I:|:|:|:|:|:|:  
Db 584 VLADSESGXITHSGM 599

RESULT 2  
US-08-868-288A-3  
Sequence 3, Application US/08868288A  
Patent No. 5922567  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Lal, Preeti  
APPLICANT: Bandman, Olga  
TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/868,288A  
FILING DATE: June 3, 1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: Pf-0309 US

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 330 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: HNT2RAT01  
CLONE: 260873  
US-08-868-288A-3

Query Match 42.6%; Score 43; DB 2; Length 330;  
Best Local Similarity 70.0%; Pred. No. 9.9;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 11 SFGAIAHGGL 20  
||||:||||  
Db 163 SFGSLGHGGL 172

## RESULT 3

US-09-235-373-3  
Sequence 3, Application US/09235373  
Patent No. 6001598

## GENERAL INFORMATION:

APPLICANT: Au-Young, Janice  
APPLICANT: Lal, Preeti  
APPLICANT: Bandman, Olga  
TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA

ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/235,373  
FILING DATE:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/868,288

FILING DATE: June 3, 1997  
ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0309 US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-845-4166  
TELEFAX: 415-855-0555

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 330 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:

LIBRARY: HNT2RAT01  
CLONE: 260873  
US-09-235-373-3

Query Match 42.6%; Score 43; DB 3; Length 330;  
Best Local Similarity 70.0%; Pred. No. 9.9;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 11 SFGAIAHGGL 20  
||||:||||  
Db 163 SFGSLGHGGL 172

## RESULT 4

US-09-388-993-3  
Sequence 3, Application US/09388993  
Patent No. 6043222

## GENERAL INFORMATION:

APPLICANT: Au-Young, Janice  
APPLICANT: Lal, Preeti  
APPLICANT: Bandman, Olga  
TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA

ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/388,993  
FILING DATE:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/868,288

FILING DATE:  
ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0309 US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-845-4166  
TELEFAX: 415-855-0555

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 330 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:

LIBRARY: HNT2RAT01  
CLONE: 260873  
US-09-388-993-3

Query Match 42.6%; Score 43; DB 3; Length 330;  
Best Local Similarity 70.0%; Pred. No. 9.9;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 11 SFGAIAHGGL 20  
||||:||||  
Db 163 SFGSLGHGGL 172

## RESULT 5

US-09-230-944-11  
Sequence 11, Application US/09230944  
Patent No. 6277380

## GENERAL INFORMATION:

APPLICANT: UEDA, Shigeharu  
APPLICANT: WATANABE, Michiko  
APPLICANT: KAWANISHI, Hitomi  
TITLE OF INVENTION: Measles Virus Mutant Antigen and Gene Coding for the

MOLECULE TYPE: protein

; GENERAL INFORMATION:

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/279,700
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/866,033A
FILING DATE: 19920408
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40399/140 NIHD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 617 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-279-700-8

Query Match          42.6%; Score 43; DB 1; Length 617;
Best Local Similarity 50.0%; Pseq. No. 20;
Matches      8; Conservative      3; Mismatches      5; Indels      0; Gaps      0

Qy      5 VMGDAESFGALHAGL 20
      1:|:|:|:|:|:|:|:|:|:
Db      584 VLADSESGHITHSGM 599

RESULT 10
US-08-279-700-10
Sequence 10, Application US/08279700
Patent No. 5578448
GENERAL INFORMATION:
APPLICANT: ROTA, Jennifer S.
APPLICANT: BELLINI, William J.
TITLE OF INVENTION: WILD-TYPE MEASLES VIRUS GLYCOPROTEINS;
TITLE OF INVENTION: VACCINE AND DETECTION METHOD THEREFOR
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
City: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/279,700
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/866,033A
FILING DATE: 19920408
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40399/140 NIHD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 10:

```



COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/348,891A  
FILING DATE: 25-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,400  
FILING DATE: 10-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 3-293625  
FILING DATE: 14-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J.  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: KP-7501  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-521-2297  
TELEFAX: 703-685-0573  
TELEX: 248425 EMBON  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 617 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-348-891A-6

Query Match 42.6%; Score 43; DB 1; Length 617;  
Best Local Similarity 50.0%; Pred. No. 20;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 VMGDAESFGAIAHGSL 20  
|:|:|:|:|:|:|:|:  
Db 584 VLADSESGHITHSGM 599

RESULT 14  
US-08-905-817-6  
Sequence 6, Application US/08905817  
Patent No. 5824777  
GENERAL INFORMATION:  
APPLICANT: SASAKI, Keiko  
APPLICANT: MORI, Takayuki  
APPLICANT: MAKINO, Satoshi  
TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,  
TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR  
NUMBER OF INVENTION: ITS ABSOLUTE IDENTIFICATION  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YOUNG & THOMPSON  
STREET: 745 South 23rd Street  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/905,817  
FILING DATE: 04-AUG-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/348,891

FILING DATE: 25-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,400  
FILING DATE: 10-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 3-293625  
FILING DATE: 14-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J.  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: KP-7501A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-521-2297  
TELEFAX: 703-685-0573  
TELEX: 248425 EMBON  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 617 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-905-817-6

Query Match 42.6%; Score 43; DB 2; Length 617;  
Best Local Similarity 50.0%; Pred. No. 20;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 VMGDAESFGAIAHGSL 20  
|:|:|:|:|:|:|:|:  
Db 584 VLADSESGHITHSGM 599

RESULT 15  
US-09-230-944-10  
Sequence 10, Application US/09230944  
Patent No. 6277380  
GENERAL INFORMATION:  
APPLICANT: UEDA, Shigeharu  
APPLICANT: WATANABE, Michiko  
APPLICANT: KAWANISHI, Hitomi  
TITLE OF INVENTION: Measles Virus Mutant Antigen and Gene Coding for the  
TITLE OF INVENTION: Same  
FILE REFERENCE: 0216-0407P  
CURRENT APPLICATION NUMBER: US/09/230,944  
EARLIER FILING DATE: 1999-02-04  
EARLIER APPLICATION NUMBER: PCT/JP98/02481  
EARLIER FILING DATE: 1998-06-04  
EARLIER APPLICATION NUMBER: 9-184285 JAPAN  
EARLIER FILING DATE: 1997-06-04  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 10  
LENGTH: 617  
TYPE: PRT  
ORGANISM: Measles virus  
FEATURE:  
OTHER INFORMATION: any n or xaa = Unknown  
US-09-230-944-10

Query Match 42.6%; Score 43; DB 4; Length 617;  
Best Local Similarity 50.0%; Pred. No. 20;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 VMGDAESFGAIAHGSL 20  
|:|:|:|:|:|:|:|:  
Db 584 VLADSESGHITHSGM 599

Search completed: July 8, 2002, 08:19:41  
Job time: 586 sec

Mon Jul 8 09:33:16 2002

us-09-582-296-2.rai

Page 8

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:20:15 ; Search time 26.89 Seconds

(without alignments)  
71.468 Million cell updates/sec

Title: US-09-582-296-2

Perfect score: 101

Sequence: 1 APAYVMGDAESFGAIAHGGL 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	47.5	749	2	S75331
2	48	47.5	750	1	H70628
3	48	47.5	763	2	T44735
4	48	47.5	767	2	H86946
5	46	45.5	121	2	F83769
6	45.5	45.0	1513	2	A70982
7	45	44.6	583	2	G81412
8	45	44.6	746	2	H84301
9	44.5	44.1	469	2	A81836
10	44.5	44.1	469	2	G81128
11	44.5	44.1	474	2	G83255
12	44.5	44.1	474	2	H82564
13	44	43.6	242	2	A11883
14	44	43.6	498	2	T37527
15	43.5	43.1	469	2	AD2909
16	43.5	43.1	469	2	AH3271
17	43.5	43.1	475	2	C97684
18	43.5	43.1	479	2	C87273
19	43.5	43.1	552	2	T07932
20	43.5	43.1	2946	2	T00867
21	43	42.6	266	2	C95950
22	43	42.6	387	2	AH1055
23	43	42.6	395	2	AB2977
24	43	42.6	422	2	A98306
25	43	42.6	617	1	HNNZHA
26	43	42.6	617	1	HNNZED
27	42.5	42.1	617	1	F84274
28	42.5	42.1	205	2	F84274
29	42	41.6	242	2	G69128

30	42	41.6	329	1	G69210	conserved hypothet
31	42	41.6	374	2	F87280	esterase A [import
32	42	41.6	618	2	A70989	hypothetical glyci
33	42	41.6	898	2	JH0440	RAD54 protein - ye
34	42	41.6	1126	2	A96032	probable two-compo
35	41.5	41.1	354	2	T46683	probable lipote p
36	41.5	41.1	358	2	H90047	hypothetical prote
37	41.5	41.1	388	2	T28693	hypothetical prote
38	41	40.6	61	2	F95014	hypothetical prote
39	41	40.6	61	2	H97887	hypothetical prote
40	41	40.6	221	2	S47690	hypothetical 25.3k
41	41	40.6	221	2	H91168	hypothetical prote
42	41	40.6	221	2	H86014	hypothetical prote
43	41	40.6	223	2	E87590	DNA-binding respon
44	41	40.6	270	2	C87568	transcription regu
45	41	40.6	323	2	A70404	acetyl-CoA carboxy

#### ALIGNMENTS

```

RESULT 1
S75331
penicillin-binding protein 1B mrcB - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein slr1710
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S75331
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-749 <KAN>
A:Cross-references: EMBL:D90904; GB:AB001339; NID:g1652225; PIRN:BA17245.1; PID:g165
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: mrcB
C:Superfamily: penicillin-binding protein

Query Match 47.5%; Score 48; DB 2; Length 749;
Best Local Similarity 37.0%; Pred. No. 13;
Matches 10; Conservative 7; Mismatches 2; Indels 8; Gaps 1;

OY 2 APAYVMGDAE-----SFGAIAHGGL 20
Db 592 PGLVVGSEARVLEMTGAYGAIANGV 618

RESULT 2
H70628
probable pkng protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: H70628
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
, Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98255987
A:Accession: H70628
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-750 <COL>
A:Cross-references: GB:Z84724; GB:AL123456; NID:g3261708; PIRN:CA06580.1; PID:g18176

```

A:Experimental source: strain H37RV  
 C:Gene: pknG  
 C:Superfamily: Mycobacterium tuberculosis probable pknG protein: protein kinase homology  
 F:149-393/Domain: protein kinase homology <KIN>

Query Match 47.5%; Score 48; DB 1; Length 750;  
 Best Local Similarity 58.8%; Pred. No. 13;  
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 4 VVMGDASFCAIAHGCL 20  
 : | | | | | | | | | |  
 Db 146 IVAGQYEVKGCIAHGCL 162

RESULT 3  
 T44735  
 serine/threonine-specific protein kinase homolog [imported] - Mycobacterium leprae  
 C:Species: Mycobacterium leprae  
 C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 04-Mar-2000  
 C:Accession: T44735  
 R:James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 Submitted to the EMBL Data Library, January 1998  
 A:Reference number: Z22831  
 A:Accession: T44735  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-763 <TAM>  
 A:Cross-references: EMBL:AL035159; PIDN:CAA22703.1  
 A:Experimental source: cosmid B1450  
 C:Genetics:  
 A:Note: MLCB1450.19c  
 C:Superfamily: Mycobacterium tuberculosis probable pknG protein: protein kinase homology

Query Match 47.5%; Score 48; DB 2; Length 763;  
 Best Local Similarity 58.8%; Pred. No. 13;  
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 4 VVMGDASFCAIAHGCL 20  
 : | | | | | | | | | |  
 Db 155 IVAGQYEVKGCIAHGCL 171

RESULT 4  
 H86946  
 Probable serine-threonine protein kinase [imported] - Mycobacterium leprae  
 C:Species: Mycobacterium leprae  
 C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
 C:Accession: H86946  
 R:Cole, S.T.; Eiglsamer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho  
 R.; Davies, R.M.; Devlin, K.; Duchoy, S.; Felwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
 eam, M.A.; Rutherford, K.M.  
 Nature 409, 1007-1011, 2001  
 A:Authors: Rutherford, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq  
 A:Title: Massive gene decay in the leprosy bacillus.  
 A:Reference number: A86909; MUID:21128732; PMID:11234002  
 A:Accession: H86946  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-767 <STO>  
 A:Cross-references: GB:AL450380; NID:g13092623; PIDN:CAC29812.1; GSPDB:GN00147  
 C:Genetics:  
 A:Gene: pknG  
 C:Superfamily: Mycobacterium tuberculosis probable pknG protein: protein kinase homology

Query Match 47.5%; Score 48; DB 2; Length 767;  
 Best Local Similarity 58.8%; Pred. No. 14;  
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 4 VVMGDASFCAIAHGCL 20

Db 159 IVAGQYEVKGCIAHGCL 175  
 : | | | | | | | | | |

RESULT 5  
 F83769  
 hypothetical protein BH0958 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: F83769  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: F83769  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-121 <STO>  
 A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA04677.1; GSPDB:G  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH0958

Query Match 45.5%; Score 46; DB 2; Length 121;  
 Best Local Similarity 56.2%; Pred. No. 3.7;  
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 5 VVMGDASFCAIAHGCL 20  
 : | | | | | | | | | |  
 Db 39 MMGDLEVFISHVAGL 54

RESULT 6  
 A70982  
 Probable ATP-dependent helicase lhr - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: A70982  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
 R.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
 Rajandream, M.A.; Rogers, J.; Rutherford, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
 A:Reference number: A70500; MUID:98295587  
 A:Accession: A70982  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1513 <COL>  
 A:Cross-references: GB:U22771; GB:AL123456; NID:g3242259; PIDN:CAB07060.1; PID: e30655  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: RV3296; lhr

Query Match 45.0%; Score 45.5; DB 2; Length 1513;  
 Best Local Similarity 45.8%; Pred. No. 73;  
 Matches 11; Conservative 3; Mismatches 5; Indels 5; Gaps 1;

OY 1 APAVVMGDASFCAIAHGCL 19  
 : | | | | | | | | | |  
 Db 316 APAHMGSSQGTGAPVLAARAHG 339

RESULT 7  
 G81412  
 aspartate--tRNA ligase (PC 6.1.1.12) Cj0640c [imported] - Campylobacter jejuni (strai  
 C:Species: Campylobacter jejuni  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 02-Jun-2000  
 C:Accession: G81412  
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chl  
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba

Nature 403, 665-668, 2000  
A>Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyf  
A:Reference number: AB1250; MUID:20150912  
A:Accession: G81412  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-563 <PAR>  
A:Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB75276.1; PID:g696810  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: asps; Cj0640c  
C:Superfamily: lysine--trna ligase  
C:Keywords: ligase

Query Match 44.6%; Score 45; DB 2; Length 583;  
Best Local Similarity 69.2%; Pred. No. 30;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 8 DAEFGAIAHGCL 20  
|| ||| |||  
DB 514 DALSFCAPHGCI 526

RESULT 8  
H84301  
hypothetical protein Vng1476c [imported] - *Halobacterium* sp. NRC-1  
C:Species: *Halobacterium* sp. NRC-1  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: H84301  
R:Ng, W.V.; Kennedy, S.P.; Mahatras, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;  
Leithner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo  
Jung, K.H.; Alm, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li  
A>Title: Genome sequence of *Halobacterium* species NRC-1.  
A:Reference number: AB4160; MUID:20504483  
A:Accession: H84301  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-746 <STO>  
A:Cross-references: GB:AE004437; NID:g10580974; PIDN:AG19780.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: VNG1476C

Query Match 44.6%; Score 45; DB 2; Length 746;  
Best Local Similarity 50.0%; Pred. No. 40;  
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 AVVMGDAESFGAIAHGCL 20  
|| : ||| : |||  
DB 220 AFEVGYITFGAIGYGI 237

RESULT 9  
AB1836  
Probable 3-isopropylmalate dehydratase (EC 4.2.1.33) large chain NMA1450 [imported] - Ne  
C:Species: *Neisseria meningitidis*  
C>Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: AB1836  
R:Parhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churche, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A>Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.  
A:Reference number: AB1775; MUID:2022556  
A:Accession: AB1836  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-469 <PAR>  
A:Cross-references: GB:AL162756; GB:AL157959; NID:g7380091; PIDN:CAB84686.1; PID:g738010  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:

A:Gene: leuc; NMA1450  
C:Superfamily: aconitate hydratase  
C:Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 44.1%; Score 44.5; DB 2; Length 469;  
Best Local Similarity 50.0%; Pred. No. 29;  
Matches 10; Conservative 3; Mismatches 2; Indels 5; Gaps 1;  
QY 4 VVMGDAE-----SFGAIAHG 18  
|| ||| : ||| |||  
DB 126 VVCGDSHTSHGAFGALAHG 145

RESULT 10  
G81128  
3-isopropylmalate dehydratase, large chain NMB1036 [imported] - *Neisseria meningitidis*  
C:Species: *Neisseria meningitidis*  
C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: G81128  
R:Teitelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.  
R.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;  
A>Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
A:Reference number: AB1000; MUID:20175755  
A:Accession: G81128  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-469 <TER>  
A:Cross-references: GB:AE002454; GB:AE002098; NID:g7226273; PIDN:AAF41435.1; PID:g722  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB1036  
C:Superfamily: aconitate hydratase

Query Match 44.1%; Score 44.5; DB 2; Length 469;  
Best Local Similarity 50.0%; Pred. No. 29;  
Matches 10; Conservative 3; Mismatches 2; Indels 5; Gaps 1;

QY 4 VVMGDAE-----SFGAIAHG 18  
|| ||| : ||| |||  
DB 126 VVCGDSHTSHGAFGALAHG 145

RESULT 11  
G83255  
3-isopropylmalate dehydratase large subunit PA3121 [imported] - *Pseudomonas aeruginosa*  
C:Species: *Pseudomonas aeruginosa*  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: G83255  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa  
A:Reference number: AB2950; MUID:20437337  
A:Accession: G83255  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-474 <STO>  
A:Cross-references: GB:AE004736; GB:AE004091; NID:g9949227; PIDN:AA06509.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: leuc; PA3121  
C:Superfamily: aconitate hydratase

Query Match 44.1%; Score 44.5; DB 2; Length 474;  
Best Local Similarity 50.0%; Pred. No. 29;  
Matches 10; Conservative 3; Mismatches 2; Indels 5; Gaps 1;



Mon Jul 8 09:33:17 2002

us-09-582-296-2.rpr

Page 5

Search completed: July 8, 2002, 08:20:16  
Job time: 576 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:21:22 ; Search time 13.48 seconds  
(without alignments)  
57.447 Million cell updates/sec

Title: US-09-582-296-2

Sequence: 1 APAYVMGDAESFGAIGHSL 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SWISSProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	47.5	750	1 PKNG_MYCTU	P96256 mycobacteri
2	48	47.5	763	1 PKNG_MYCLE	P57993 mycobacteri
3	44.5	44.1	323	1 AOP4_BOVIN	077750 bos taurus
4	44.5	44.1	323	1 AOP4_HUMAN	P55087 homo sapien
5	44	43.6	498	1 YAOI_SCHPO	Q10097 schizosach
6	43	42.6	326	1 DJB6_HUMAN	075190 homo sapien
7	43	42.6	617	1 HEMA_MEASA	P35971 measles vir
8	43	42.6	617	1 HEMA_MEASH	P08362 measles vir
9	43	42.6	617	1 HEMA_MEASH	P66083 measles vir
10	43	42.6	617	1 HEMA_MEASH	P26008 measles vir
11	42	41.6	317	1 O2FL_HUMAN	Q13607 homo sapien
12	42	41.6	329	1 Y830_MERTH	Q26918 hominobact
13	42	41.6	898	1 RAS4_YEAST	P32863 saccharomyc
14	41	40.6	185	1 APT_CORGL	087330 coxyriabact
15	41	40.6	221	1 YHHQ_ECOLI	P37619 escherichia
16	41	40.6	323	1 ACCA_AQJAE	067260 aquifex aeo
17	41	40.6	465	1 LUDH_CHLTR	084561 chlamydia t
18	40.5	40.1	357	1 YNS7_YEAST	P53873 saccharomyc
19	40.5	40.1	424	1 SYH_BACSU	032039 bacillus su
20	40.5	40.1	968	1 Y682_MERVA	P58095 methanococ
21	40	39.6	282	1 YOR3_MERVA	P68000 methanococ
22	40	39.6	317	1 O2F2_HUMAN	095006 homo sapien
23	40	39.6	591	1 SYD_PSEAE	051422 pseudomonas
24	39.5	39.1	314	1 VANB_PSEAS	P12380 pseudomonas
25	39.5	39.1	448	1 GCSI_BACSU	P54376 bacillus su
26	39.5	39.1	661	1 Y182_SYNV3	055774 synechocyst
27	39	38.6	61	1 AERC_AERSO	P09165 aeromonas s
28	39	38.6	151	1 RS9_MYCTU	006259 mycobacteri
29	39	38.6	242	1 DJB6_MOUSE	054946 mus musculu
30	39	38.6	291	1 ISPA_MICU	066126 mlctococcus
31	39	38.6	328	1 NTR3_AZOBR	P45672 azospirillum
32	39	38.6	342	1 EFTS_LACIA	09cd55 lactococcus
33	39	38.6	345	1 EFTS_STRPN	P80715 streptococ

34	39	38.6	426	1 GABT_ECOLI	P22256 escherichia
35	39	38.6	453	1 VG_DROME	Q26366 drosophila
36	39	38.6	480	1 CALB_PSESP	086447 pseudomonas
37	39	38.6	481	1 GLNA_HELPY	P94845 helicobacte
38	39	38.6	484	1 US15_HCMVA	P09718 human cytom
39	39	38.6	528	1 GD_DROME	062589 drosophila
40	39	38.6	587	1 RECN_MYCTU	033197 mycobacteri
41	39	38.6	1051	1 UBAL_WHEAT	P20973 triticum ae
42	39	38.6	1051	1 UBA2_WHEAT	P31251 triticum ae
43	39	38.6	1091	1 DIA_DROME	P48608 drosophila
44	39	38.6	1197	1 EVGS_ECOLI	P58402 escherichia
45	39	38.6	1197	1 EVGS_ECOLI	P30855 escherichia

## ALIGNMENTS

RESULT	ID	PKNG_MYCTU	STANDARD	PRT	750 AA.
1	PKNG_MYCTU				
AC	P96256				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Probable serine/threonine-protein kinase pknG (EC 2.7.1.-).				
GN	PKNG OR RV0410C OR MT0423 OR MTCY22G10.06C.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=H37RV;				
RX	MEDLINE=98295967; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eigler K., Gas S., Barry C.E. III, Tekala F.,				
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,				
RA	Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutter S., Seeger K., Skellon S., Squares S., Squares R.,				
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;				
RT	Deciphering the biology of Mycobacterium tuberculosis from the				
RT	complete genome sequence."				
RL	Nature 393:537-544(1998).				
[2]					
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CDC 1551 / Oshkosh;				
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,				
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,				
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,				
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,				
RT	Bisai W.;				
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and				
RT	laboratory strains."				
RT	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.				
RL	-1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.				
CC					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sdb.ch">license@isb-sdb.ch</a> ).				
CC					
DR	EMBL: Z44724; CAB0580.1; -				
DR	EMBL: AE006946; AAK44647.1; ALT_INT.				
DR	TIGR: MT0423; -				
DR	Tuberculist: RV0410c; -				
DR	InterPro: IPR000719; Euk_pkinase.				
DR	InterPro: IPR002290; Ser_thr_pkinase.				
DR	Pfam: PF00069; pkinase; 1.				

DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE: PS00111; PROTEIN\_KINASE\_DOM; 1.  
 KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;  
 KM ATP-binding; Complete proteome.  
 FT DOMAIN 151 396  
 FT NP\_BIND 157 165 ATP (BY SIMILARITY).  
 FT BINDING 181 181 ATP (BY SIMILARITY).  
 FT ACT\_SITE 276 276 BY SIMILARITY.  
 SQ SEQUENCE 750 AA; 81577 MW; CFC6E54DBE19569F CRC64;

Query Match 47.5%; Score 48; DB 1; Length 750;  
 Best Local Similarity 58.8%; Pred. No. 6.6;  
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 VVWGDAEFGAIAHGCL 20  
 Db 146 IVAGQYEVWGCIAGHGL 162

RESULT 2  
 PKNG\_MYCLE STANDARD; PRT; 763 AA.  
 AC P57993; Q9ZBL8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Probable serine/threonine-protein kinase phnd (EC 2.7.1.-).  
 GN PKNG OR ML0304 OR MLCB1450.19C.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies K.M., Devlin K., Dutoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagsels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skellern J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrall B.G.,  
 RT "Massive gene decay in the leprosy bacillus."  
 RL Nature 409:1007-1011(2001).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AL035159; CA22703.1; -;  
 DR EMBL: AL583918; CAC29812.1; ALT\_INIT.  
 DR Leptoma; ML0304;  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR SMART; SM00220; S\_TKC; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE: PS00111; PROTEIN\_KINASE\_DOM; 1.  
 KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;  
 KM ATP-binding; Complete proteome.

FT DOMAIN 160 406 PROTEIN\_KINASE.  
 FT NP\_BIND 166 174 ATP (BY SIMILARITY).  
 FT BINDING 190 190 ATP (BY SIMILARITY).  
 FT ACT\_SITE 289 289 BY SIMILARITY.  
 SQ SEQUENCE 763 AA; 83469 MW; 9B42725EB49C1199 CRC64;

Query Match 47.5%; Score 48; DB 1; Length 763;  
 Best Local Similarity 58.8%; Pred. No. 6.7;  
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 VVWGDAEFGAIAHGCL 20  
 Db 155 IVAGQYEVWGCIAGHGL 171

RESULT 3  
 AQP4\_BOVIN STANDARD; PRT; 323 AA.  
 AC O77750;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, last sequence update)  
 DT 01-MAR-2002 (Rel. 41, last annotation update)  
 DE Aquaporin 4 (WCH4) (Mercurial-insensitive water channel) (MWC).  
 GN AQP4 OR AQP-4.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC STRAIN=Holstein; TISSUE=Brain;  
 RX PubMed=10673041;  
 RA Sobue K., Yamamoto N., Yoneda K., Fujita K., Miura Y., Asai K.,  
 RA Tsuda T., Katsuya H., Kato T.;  
 RT "Molecular cloning of two bovine aquaporin-4 cDNA isoforms and their  
 RT expression in brain endothelial cells."  
 RL Biochim. Biophys. Acta 1489:393-398(1999).  
 CC -1- FUNCTION: FORMS A WATER-SPECIFIC CHANNEL. OSMORECEPTOR WHICH  
 CC REGULATES BODY WATER BALANCE AND MEDIATES WATER FLOW WITHIN THE  
 CC CENTRAL NERVOUS SYSTEM.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1/A and 2/B (shown here); are  
 CC produced by alternative splicing.  
 CC -1- SIMILARITY: BELONGS TO THE MIP/AQUAPORIN FAMILY (TC 1.A.8).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB015947; BAA36505.2; -;  
 DR EMBL: AB012950; BAA33583.1; -;  
 DR EMBL: AB028642; BAA89291.1; -;  
 DR HSSP; P29972; 1RY.  
 DR InterPro: IPR000425; MIP.  
 DR Pfam; PF00230; MIP; 1.  
 DR PRINTS; PR00783; MINTRINSICP.  
 DR PROSITE: PS00221; MIP; 1.  
 KM Transport; Transmembrane; Alternative splicing.  
 FT DOMAIN 1 36  
 FT TRANSMEM 1 36 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 37 57 POTENTIAL.  
 FT DOMAIN 58 64 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 65 85 POTENTIAL.  
 FT DOMAIN 86 115 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 116 136 POTENTIAL.  
 FT DOMAIN 137 155 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 156 176 POTENTIAL.  
 FT DOMAIN 177 184 CYTOPLASMIC (POTENTIAL).

CC TRANSMEM 185 205 POTENTIAL.  
 CC FT DOMAIN 206 231 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 232 252 POTENTIAL.  
 CC FT DOMAIN 253 323 CYTOPLASMIC (POTENTIAL).  
 CC FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT VARSPLIC 1 22 MISSING (IN ISOFORM 1).  
 CC SQ SEQUENCE 323 AA: 34616 MW: E8E14507E647AD5F CRC64;

Query Match 44.1%; Score 44.5; DB 1; Length 323;  
 Best Local Similarity 39.3%; Pred. No. 10;  
 Matches 11; Conservative 4; Mismatches 4; Indels 9; Gaps 1;

OY 2 PAVVMGDAES-----FGAIAHGGL 20  
 Db 220 PAVINGMNMENHWYVGPITGAVLAGL 247

RESULT 4  
 ID AQP4\_HUMAN STANDARD: PRT: 323 AA.  
 AC P55087; P78564; Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Aquaporin 4 (MCH4) (Mercurial-insensitive water channel) (MIMC).  
 GN AQP4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC TISSUE=Petal brain;  
 RX MEDLINE=96032721; PubMed=7559426;  
 RA Yang B., Ma T., Verkman A.S.;  
 RT "CDNA cloning, gene organization, and chromosomal localization of a  
 human mercurial insensitive water channel. Evidence for distinct  
 RT transcriptional units.";  
 RT J. Biol. Chem. 270:22907-22913(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96176324; PubMed=8601457;  
 RA Miska T., Abe K., Iwabuchi K., Kusakabe Y., Ichinose M., Miki K.,  
 RA Emori Y., Arai S.;  
 RT "A water channel closely related to rat brain aquaporin 4 is  
 RT expressed in acid- and pepsinogen-secretory cells of human stomach.";  
 RT FEBS Lett. 381:208-212(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97008105; PubMed=8655281;  
 RA Lu M., Lee M.D., Smith B.L., Jung J.S., Agre P., Verdi J.K. M.A.J.,  
 RA Merx G., Rjiss J.P.L., Deen P.M.T.;  
 RT "The human AQP4 gene: definition of the locus encoding two water  
 RT channel polypeptides in brain.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 93:10908-10912(1996).  
 CC -1- FUNCTION: FORMS A WATER-SPECIFIC CHANNEL. OSMORECEPTOR WHICH  
 CC REGULATES BODY WATER BALANCE AND MEDIATES WATER FLOW WITHIN THE  
 CC CENTRAL NERVOUS SYSTEM.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 AND 2 (SHOWN HERE); ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: BRAIN - MUSCLE >> HEART, KIDNEY, LUNG, AND  
 CC TRACHEA.  
 CC -1- SIMILARITY: BELONGS TO THE MIP/AQUAPORIN FAMILY (TC 1.A.8).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 CC EMBL: U34846; AAC52112.1; ALT\_INT.  
 CC EMBL: U34845; AAC50284.1; ALT\_INT.  
 CC EMBL: D63412; BAA09715.1; -.  
 CC EMBL: U63622; AAB26957.1; -.  
 CC EMBL: U63623; AAB26958.1; -.  
 CC HSSP: P29972; IFQY.  
 CC MIM: 600308; -.  
 CC InterPro: IPR000425; MIP.  
 CC Pfam: PF00230; MIP; 1.  
 CC PRINTS: PR00783; MINTRINSICP.  
 CC PROSITE: PS00221; MIP; 1.  
 CC Transprot; Transmembrane; Alternative splicing.  
 CC FT DOMAIN 1 36 CYTOPLASMIC (POTENTIAL).  
 CC FT TRANSMEM 37 57 POTENTIAL.  
 CC FT DOMAIN 58 64 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 65 85 POTENTIAL.  
 CC FT DOMAIN 86 115 CYTOPLASMIC (POTENTIAL).  
 CC FT TRANSMEM 116 136 POTENTIAL.  
 CC FT DOMAIN 137 155 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 156 176 POTENTIAL.  
 CC FT DOMAIN 177 184 CYTOPLASMIC (POTENTIAL).  
 CC FT TRANSMEM 185 205 POTENTIAL.  
 CC FT DOMAIN 206 231 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 232 252 POTENTIAL.  
 CC FT DOMAIN 253 323 CYTOPLASMIC (POTENTIAL).  
 CC FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT VARSPLIC 1 22 MISSING (IN ISOFORM 1).  
 CC FT CONFLICT 246 246 G -> A (IN REF. 1).  
 CC FT CONFLICT 287 288 VE -> K (IN REF. 1).  
 CC FT CONFLICT 296 296 P -> L (IN REF. 1).  
 CC SQ SEQUENCE 323 AA: 34829 MW: 1A1600CF0DC11052 CRC64;

Query Match 44.1%; Score 44.5; DB 1; Length 323;  
 Best Local Similarity 39.3%; Pred. No. 10;  
 Matches 11; Conservative 4; Mismatches 4; Indels 9; Gaps 1;

OY 2 PAVVMGDAES-----FGAIAHGGL 20  
 Db 220 PAVINGMNMENHWYVGPITGAVLAGL 247

RESULT 5  
 ID YAOI\_SCHPO STANDARD: PRT: 498 AA.  
 AC O10097;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Putative transporter C11D3.18C.  
 GN SPAC11D3.18C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 CC OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Connor R., Churcher C.M., Barrett B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE ATLANTOITE PERMEASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)



CC Paramyxoviridae; Paramyxovirinae; Morbillivirus.  
RX NCBI\_TaxID=36408;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=93227570; PubMed=8470368;  
RT Mori T., Sasaki K., Hashimoto H., Makino S.;  
RT "Molecular cloning and complete nucleotide sequence of genomic RNA of  
RT the A/K-C strain of attenuated measles virus.";  
RT Virus Genes 7:67-81(1993).  
RL  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS  
CC TO CELL RECEPTORS AND FOR INITIATING INFECTION. NEURAMINIDASE  
CC ACTIVITY HELPS THE EFFICIENT SPREAD OF THE VIRUS BY DISSOCIATING  
CC THE MATURE VIRIONS FROM THE NEURAMINIC ACID CONTAINING  
CC GLYCOPROTEINS.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,  
CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in  
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and  
CC synthetic substrates.  
CC -1- SUBCELLULAR LOCATION: EXTERNAL, ANCHORED TO THE ENVELOPE BY ITS  
CC N-TERMINAL HYDROPHOBIC SEQUENCE.  
CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES HEMAGGLUTININ-  
CC NEURAMINIDASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: S58435; AAB26146.1; -  
DR PIR: F48556; F48556.  
DR InterPro: IPR000665; Hem-neuramidase.  
DR Pfam: PF00423; HN; 1.  
KW Hydrolyase; Hemagglutinin; Envelope protein; Glycoprotein;  
KW Transmembrane.  
KM  
FT DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 35 58 MEMBRANE ANCHOR (POTENTIAL).  
FT DOMAIN 59 617 EXTRACELLULAR (POTENTIAL).  
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 617 AA; 69322 MW; F7D48744436C1CA CRC64;  
Query Match 42.68; Score 43; DB 1; Length 617;  
Best Local Similarity 50.08; Pred. No. 33;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
QY 5 VMGDAESFGAIAHGGL 20  
Db 584 VLADSESGHITSGM 599  
RESULT 8  
HEMA\_MEASH STANDARD; PRT; 617 AA.  
ID HEMA\_MEASH  
AC P08362;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin-neuraminidase (EC 3.2.1.18).  
GN HN.  
OS Measles virus (strain Edmonston) (Subacute sclerosing panencephalitis  
OS virus).  
CC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
CC Paramyxoviridae; Paramyxovirinae; Morbillivirus.  
OX NCBI\_TaxID=11235;  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=86181590; PubMed=3008420;  
RA Alkhatib G., Brisdie D.J.;  
RT "The predicted primary structure of the measles virus hemagglutinin.";  
RT Virology 150:479-490(1986).  
RL  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=90085790; PubMed=2596022;  
RT Cattaneo R., Schmidt A., Spielhofer P., Kaelin K., Bacsko K.,  
RT Meulen V., Pardowitz J., Flanagan S., Rima B.K., Udem S.A.;  
RT "Mutated and hypermutated genes of persistent measles viruses which  
RT caused lethal human brain diseases.";  
RL Virology 173:415-425(1989).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS  
CC TO CELL RECEPTORS AND FOR INITIATING INFECTION. NEURAMINIDASE  
CC ACTIVITY HELPS THE EFFICIENT SPREAD OF THE VIRUS BY DISSOCIATING  
CC THE MATURE VIRIONS FROM THE NEURAMINIC ACID CONTAINING  
CC GLYCOPROTEINS.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,  
CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in  
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and  
CC synthetic substrates.  
CC -1- SUBCELLULAR LOCATION: EXTERNAL, ANCHORED TO THE ENVELOPE BY ITS  
CC N-TERMINAL HYDROPHOBIC SEQUENCE.  
CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES HEMAGGLUTININ-  
CC NEURAMINIDASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: M14877; AAA46424.1; -  
DR EMBL: K01711; AAA75500.1; -  
DR PIR: A27006; HNMZED.  
DR InterPro: IPR000665; Hem-neuramidase.  
DR Pfam: PF00423; HN; 1.  
KW Hydrolyase; Hemagglutinin; Envelope protein; Glycoprotein;  
KW Transmembrane.  
KM  
FT DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 35 58 MEMBRANE ANCHOR (POTENTIAL).  
FT DOMAIN 59 617 EXTRACELLULAR (POTENTIAL).  
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 617 AA; 69249 MW; 0E5A05AEDA43D9C6 CRC64;  
Query Match 42.68; Score 43; DB 1; Length 617;  
Best Local Similarity 50.08; Pred. No. 33;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
QY 5 VMGDAESFGAIAHGGL 20  
Db 584 VLADSESGHITSGM 599  
RESULT 9  
HEMA\_MEASH STANDARD; PRT; 617 AA.  
ID HEMA\_MEASH  
AC P06830;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin-neuraminidase (EC 3.2.1.18).  
GN HN.  
OS Measles virus (strain Halle) (Subacute sclerosing panencephalitis  
OS virus).  
CC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
CC Paramyxoviridae; Paramyxovirinae; Morbillivirus.

CC Paramyxoviridae; Paramyxovirinae; Morbillivirus.  
OX NCBI\_TaxID=11236;  
[1]  
RX SEQUENCE FROM N.A.  
RA MEDLINE=87085487; PubMed=3794664;  
RT "Gerald C., Buckland R., Barker R., Freeman G., Wild T.F.;  
RT "Measles virus hemagglutinin gene: cloning, complete nucleotide  
RT sequence analysis and expression in COS cells.";  
RL J. Gen. Virol. 67:2695-2703(1986).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS  
CC TO CELL RECEPTORS AND FOR INITIATING INFECTION. NEURAMINIDASE  
CC ACTIVITY HELPS THE EFFICIENT SPREAD OF THE VIRUS BY DISSOCIATING  
CC THE MATURE VIRIONS FROM THE NEURAMINIC ACID CONTAINING  
CC GLYCOPROTEINS.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,  
CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in  
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and  
CC synthetic substrates.  
CC -1- SUBCELLULAR LOCATION: EXTERNAL, ANCHORED TO THE ENVELOPE BY ITS  
CC N-TERMINAL HYDROPHOBIC SEQUENCE.  
CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES HEMAGGLUTININ-  
CC NEURAMINIDASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X04720; CAA28427.1; -  
DR PIR: A27007; HMNZHA.  
DR InterPro: IPR000665; Hem-neuramndse.  
DR Pfam: PF00423; HN; 1.  
KW Hydrolase; Hemagglutinin; Envelope protein; Glycoprotein;  
KW Transmembrane.  
FT DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 35 58 MEMBRANE ANCHOR (POTENTIAL).  
FT DOMAIN 59 617 EXTRACELLULAR (POTENTIAL).  
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 617 AA; 69077 MW; 39240D495A525C72 CRC64;  
Query Match 42.6%; Score 43; DB 1; Length 617;  
Best Local Similarity 50.0%; Pred. No. 33;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
QY 5 VMGDAESFGAIAHGGL 20  
Db 584 VLADSESGHITHSGM 599  
RESULT 10  
HEMA\_MEASI STANDARD; PRT; 617 AA.  
ID HEMA\_MEASI STANDARD; PRT; 617 AA.  
AC P26028;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin-neuraminidase (EC 3.2.1.18).  
GN HN.  
OS Measles virus (strain IP-3-Ca) (Subacute sclerosing panencephalitis  
OS virus).  
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
OC Paramyxoviridae; Paramyxovirinae; Morbillivirus.  
OX NCBI\_TaxID=11237;  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=92263801; PubMed=1585658;  
RA Schmidt A., Spielhofer P., Cattaneo R., Bacsko K., Ter Meulen V.,  
RA Billeter M.A.;  
RT "Subacute sclerosing panencephalitis is typically characterized by  
RT alterations in the fusion protein cytoplasmic domain of the  
RT persisting measles virus.";  
RL Virology 188:910-915(1992).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS  
CC TO CELL RECEPTORS AND FOR INITIATING INFECTION. NEURAMINIDASE  
CC ACTIVITY HELPS THE EFFICIENT SPREAD OF THE VIRUS BY DISSOCIATING  
CC THE MATURE VIRIONS FROM THE NEURAMINIC ACID CONTAINING  
CC GLYCOPROTEINS.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,  
CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in  
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and  
CC synthetic substrates.  
CC -1- SUBCELLULAR LOCATION: EXTERNAL, ANCHORED TO THE ENVELOPE BY ITS  
CC N-TERMINAL HYDROPHOBIC SEQUENCE.  
CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES HEMAGGLUTININ-  
CC NEURAMINIDASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X16566; CAA34569.1; -  
DR InterPro: IPR000665; Hem-neuramndse.  
DR Pfam: PF00423; HN; 1.  
KW Hydrolase; Hemagglutinin; Envelope protein; Glycoprotein;  
KW Transmembrane.  
FT DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 35 58 MEMBRANE ANCHOR (POTENTIAL).  
FT DOMAIN 59 617 EXTRACELLULAR (POTENTIAL).  
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 617 AA; 66607 MW; 8F518CE44ACB4F3 CRC64;  
Query Match 42.6%; Score 43; DB 1; Length 617;  
Best Local Similarity 50.0%; Pred. No. 33;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
QY 5 VMGDAESFGAIAHGGL 20  
Db 584 VLADSESGHITHSGM 599  
RESULT 11  
O2FL\_HUMAN STANDARD; PRT; 317 AA.  
ID O2FL\_HUMAN STANDARD; PRT; 317 AA.  
AC O13607; O9UDX1;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Olfactory receptor 2P1 (Olfactory receptor-like protein OLF3).  
GN OR2F1 OR OLF3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Issel-Tarver L., Rine J.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBD databases.  
RN [2]  
RP SEQUENCE FROM N.A.



DR EMBL: Z48618; CAA88534.1; -  
 DR EMBL: Z72685; CAA96875.1; -  
 DR PIR: JH0440; JH0440.  
 DR SGD: S0003131; RAD54.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR000330; SNF\_N.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00176; SNF2\_N; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR SMART: SM00490; HELIC; 1.  
 DR SMART: SM00490; HELIC; 1.  
 DR DNA repair: Nuclear protein; DNA-binding; Helicase; ATP-binding.  
 FT DOMAIN 21 25 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT NE\_BIND 335 342 ATP (POTENTIAL).  
 FT SITE 455 458 DECH BOX.  
 FT SEQUENCE 898 AA: 101754 MW: 26C0A9B7E3D66292 CRC64;

Query Match 41.6%; Score 42; DB 1; Length 898;  
 Best Local Similarity 40.0%; Pred. No. 67;  
 Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 APANVMDAESFGAIAHGL 20  
 DB 853 APANVMDAESFGAIAHGL 872

RESULT 14  
 APTL\_CONGL STANDARD; PRT; 185 AA.  
 AC 087330;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT).  
 GN APT.  
 OS Corynebacterium glutamicum (Brevibacterium flavum).  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
 OC Corynebacterium.  
 OX NCBI\_TaxID=1718;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032;  
 RX MEDLINE=98361031; PubMed=9695918;  
 RA Wehner L., Schaefer A., Burkowski A., Kraemer R., Mechold U.,  
 RA Malke H., Puchner A., Kalinowski J.;  
 RT "The role of the Corynebacterium glutamicum rel gene in (p)ppGpp  
 metabolism.";  
 RL Microbiology 144:1853-1862(1998).  
 CC -1- FUNCTION: CATALYZES A SALVAGE REACTION RESULTING IN THE FORMATION  
 CC OF AMP, THAT IS ENERGETICALLY LESS COSTLY THAN DE NOVO SYNTHESIS.  
 CC -1- CATALYTIC ACTIVITY: AMP + diphosphate = adenine + 5-phospho-  
 CC alpha-D-ribose 1-diphosphate.  
 CC -1- PATHWAY: PURINE SALVAGE.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE  
 CC PHOSPHORIBOSYLTRANSFERASE FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: AF038651; AAC35493.1; -  
 DR InterPro: IPR000836; Priboyltran.  
 DR InterPro: IPR002375; Pur\_pyr\_pr\_transf.  
 DR Pfam: PF00156; Priboyltran; 1.

DR PROSITE: PS00103; PUR\_PYR\_PR\_TRANSFER; 1.  
 KW Transferase; Glycosyltransferase; Purine salvage.  
 SQ SEQUENCE 185 AA: 19563 MW: FDBAC03A3FF18182 CRC64;

Query Match 40.6%; Score 41; DB 1; Length 185;  
 Best Local Similarity 72.7%; Pred. No. 21;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 VMDAESFGAI 15  
 DB 39 VMDAESFGAV 49

RESULT 15  
 YHQ\_ECOLI STANDARD; PRT; 221 AA.  
 ID YHQ\_ECOLI  
 AC P37619;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein yhq.  
 GN YHQ OR B3471.  
 GN YHQ OR B3471.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=94316500; PubMed=8041620;  
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
 RT region from 76.0 to 81.5 minutes.";  
 RL Nucleic Acids Res. 22:2576-2586(1994).  
 CC -1- SIMILARITY: STRONG, TO H. INFLUENZAE H10862.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: U00039; AAB18446.1; -  
 DR EMBL: AE000423; AAC76496.1; -  
 DR Ecogene: EG1217; yhq.  
 DR InterPro: IPR003744; DUF165.  
 DR Pfam: PF02592; DUF165; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 221 AA: 25269 MW: E3D90237BEE7ABCC CRC64;

Query Match 40.6%; Score 41; DB 1; Length 221;  
 Best Local Similarity 53.3%; Pred. No. 25;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 6 MDAESFGAIAHGL 20  
 DB 92 MGSWGFALAHFNL 106

Search completed: July 8, 2002, 08:21:24  
 Job time: 184 sec

Mon Jul 8 09:33:18 2002

us-09-582-296-2.rsp

Page 9



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:21:03 ; Search time 41.68 Seconds  
(without alignments)  
83.011 Million cell updates/sec

Title: US-09-582-296-2  
Perfect score: 101  
Sequence: 1 APAYVWGDAESFGAIHAGGL 20

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPREMBL\_19:\*
- 2: sp.archaea:\*
- 3: sp.bacteria:\*
- 4: sp.fungi:\*
- 5: sp.human:\*
- 6: sp.invertebrate:\*
- 7: sp.mammal:\*
- 8: sp.mhc:\*
- 9: sp.phage:\*
- 10: sp.plant:\*
- 11: sp.potent:\*
- 12: sp.virus:\*
- 13: sp.vertibrate:\*
- 14: sp.unclassified:\*
- 15: sp.rvivirus:\*
- 16: sp.bacteriaph:\*
- 17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	47.5	749	16 P73218	P73218 synechocyst
2	48	47.5	903	2 Q91255	Q91255 streptomyc
3	48	47.5	936	2 Q9AMT5	Q9AMT5 bradyrhizob
4	47	46.5	474	12 Q89542	Q89542 bovine herp
5	46	45.5	105	2 Q929R6	Q929R6 bacillus ha
6	46	45.5	121	16 Q9KE98	Q9KE98 bacillus ha
7	46	45.5	273	16 Q98106	Q98106 rhizobium I
8	45.5	45.0	899	5 Q9VWQ0	Q9VWQ0 drosophila
9	45.5	45.0	1513	16 P96901	P96901 mycobacteri
10	45	44.6	583	16 Q9PHM7	Q9PHM7 campylobact
11	45	44.6	746	17 Q9HPT7	Q9HPT7 halobacteri
12	44.5	44.1	178	6 Q46426	Q46426 oryctolagus
13	44.5	44.1	302	11 Q923U5	Q923U5 dipodomys m
14	44.5	44.1	324	11 Q923U4	Q923U4 dipodomys m
15	44.5	44.1	469	16 Q9J215	Q9J215 neisseria m
16	44.5	44.1	469	16 Q9J082	Q9J082 neisseria m

17	44.5	44.1	474	16 Q9PAX0	Q9PAX0 xylella fas
18	44.5	44.1	474	16 Q9HZA3	Q9HZA3 pseudomonas
19	44	43.6	610	12 Q9WJ08	Q9WJ08 measles vir
20	43.5	43.1	448	2 Q9EV53	Q9EV53 rhizobium m
21	43.5	43.1	469	16 Q98EP1	Q98EP1 rhizobium l
22	43.5	43.1	469	16 Q92L76	Q92L76 rhizobium m
23	43.5	43.1	479	16 Q9ABN0	Q9ABN0 caulobacter
24	43.5	43.1	552	10 P93837	P93837 brasica na
25	43.5	43.1	2946	10 Q64634	Q64634 arabidopsis
26	43	42.6	187	5 Q45045	Q45045 scitropnaga
27	43	42.6	244	16 Q9YJ05	Q9YJ05 streptococ
28	43	42.6	266	16 Q92W38	Q92W38 rhizobium m
29	43	42.6	407	12 Q98376	Q98376 measles vir
30	43	42.6	407	12 Q98378	Q98378 measles vir
31	43	42.6	407	12 Q98379	Q98379 measles vir
32	43	42.6	407	12 Q98380	Q98380 measles vir
33	43	42.6	407	12 Q98381	Q98381 measles vir
34	43	42.6	407	12 Q98382	Q98382 measles vir
35	43	42.6	407	12 Q98383	Q98383 measles vir
36	43	42.6	407	12 Q98384	Q98384 measles vir
37	43	42.6	407	12 Q98385	Q98385 measles vir
38	43	42.6	407	12 Q98386	Q98386 measles vir
39	43	42.6	407	12 Q98387	Q98387 measles vir
40	43	42.6	407	12 Q98388	Q98388 measles vir
41	43	42.6	407	12 Q98389	Q98389 measles vir
42	43	42.6	407	12 Q98390	Q98390 measles vir
43	43	42.6	407	12 Q98392	Q98392 measles vir
44	43	42.6	407	12 Q98393	Q98393 measles vir
45	43	42.6	407	12 Q98394	Q98394 measles vir

## ALIGNMENTS

Query Match	Score	DB	Length
47.5%	Score 48;	DB 16;	Length 749;



DR EMBL: AJ004801; CAA06095.1; -  
SQ SEQUENCE 474 AA; 49975 MW; 63C977A0BADE6AC7 CRC64;

Query Match 46.5%; Score 47; DB 12; Length 474;  
Best Local Similarity 60.0%; Pred. No. 37;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 6 MGDASFCAIAGGL 20  
:|||||:|  
DB 173 LGDASSYGAVPDGL 187

RESULT 5  
Q929R6 PRELIMINARY; PRT; 105 AA.  
AC Q929R6; 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DE 01-MAY-2001 (TREMBLrel. 16, Last annotation update)  
DE UNKNOWN (FRAGMENT).  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125;  
RX MEDLINE=99184645; PubMed=10086841;  
RA Takami H., Nakasone K., Hirama C., Takaki Y., Masui N., Fujii F.,  
Nakamura Y., Inoue A.;  
RT "An improved physical and genetic map of the genome of alkaliphilic  
Bacillus sp. C-125".  
RL Extremophiles 3:21-28(1999).  
DR EMBL: AB013366; BAA73366.1; -.  
FT NON\_TER 105  
SQ SEQUENCE 105 AA; 11853 MW; 7E18ACB7566A2F88 CRC64;

Query Match 45.5%; Score 46; DB 2; Length 105;  
Best Local Similarity 56.2%; Pred. No. 9.5;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 5 VMGDAESFGAIAHGL 20  
:|||||:|  
DB 39 MGDLEVFHSHVHAGL 54

RESULT 6  
Q9KE98 PRELIMINARY; PRT; 121 AA.  
AC Q9KE98; 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE BH0958 PROTEIN.  
GN BH0958.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512382; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
halodurans and genomic sequence comparison with Bacillus subtilis".  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL: AP001510; BAB04677.1; -.  
KW Complete proteome.

SQ SEQUENCE 121 AA; 13594 MW; ABE59AF3911A4F7B CRC64;

Query Match 45.5%; Score 46; DB 16; Length 121;  
Best Local Similarity 56.2%; Pred. No. 11;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 5 VMGDAESFGAIAHGL 20  
:|||||:|  
DB 39 MGDLEVFHSHVHAGL 54

RESULT 7  
Q98L06 PRELIMINARY; PRT; 273 AA.  
AC Q98L06; 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE MLR1236 PROTEIN.  
GN MLR1236.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
Watanabe A., Idekawa K., Ishikawa A., Kawashima K., Kimura T.,  
Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
Takuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
Mesorhizobium loti".  
RL DNA Res. 7:331-338(2000).  
DR EMBL: AP002996; BAB48657.1; -.  
DR InterPro: IPR002376; formyl\_transf.  
DR Pfam: PF00551; formyl\_transf. 1.  
KW Complete proteome.  
SQ SEQUENCE 273 AA; 29214 MW; 6975846D61F5E2D2 CRC64;

Query Match 45.5%; Score 46; DB 16; Length 273;  
Best Local Similarity 53.3%; Pred. No. 28;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 AVVMGDAESFGAIAH 17  
:|:|:|:|:|  
DB 177 ALVSGDAQNFQTTVH 191

RESULT 8  
Q9VNO0 PRELIMINARY; PRT; 899 AA.  
AC Q9VNO0; 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE CG10712 PROTEIN.  
GN CG10712 OR CG18622 OR CG18623.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydriidae; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

Query Match	45.08;	Score 45.5;	DB 5;	Length 899;
Best Local Similarity	55.08;	Pred. No. 1.3e+02;		
Matches 11;	Conservative 2;	Mismatches 6;	Indels 1;	Gaps 1

Query Match	Best Local Similarity	45.0%;	Score 45.5;	DB 16;	Length 1513;
Matches 11;	Conservative 3;	Mismatches 5;	Indels 5;	Gaps 1;	
1 APAYWGDASFSG-----AIAHCG 19	111 :11 :::11 :11 :111				
316 APAHIMSGGTCFGCAPVLARAHHG 339					
RA Gordon V., Eglmeier K., Gas S., Barry C.E. III, Tekla F., RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R., RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S., RA Holsen T., Jagers K., Krogh A., McLean J., Moule S., Murphy L., RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., RA Ruter S., Seeger K., Skellon S., Squares S., Squares R., RA Sauton J.E., Taylor K., Whitehead S., Barrall B.G.; RT "pigeonholing the biology of Mycobacterium tuberculosis from the RT complete genome sequence." RL Nature 393:537-544(1998). CC -1- SIMILARITY: TO DEAD/DEAH BOX HELICASE FAMILY. CC -1- SIMILARITY: TO HELICASE C-TERMINAL DOMAIN. DR EMBL: 292771; CAB07060.1; - DR Tuberculist; RV3296; - DR InterPro: IPR002106; AA_TRNA_Ligase_II. DR InterPro: IPR001410; DEAD. DR InterPro: IPR001650; Helicase_C. DR Pfam; PF00270; DEAD; 1. DR Pfam; PF00271; helicase_C; 1. DR SMART; SM00487; DEXDC; 1. DR SMART; SM00490; HELIC; 1. DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1. KM ATP-binding; Complete proteome; Helicase. KW SEQUENCE 1513 AA; 161378 MW; 6E96F89D83066B52 CRC64;					

DR InterPro: IPR002313; tRNA-synt\_Lys\_2.  
 DR InterPro: IPR004365; tRNA\_antl.  
 DR Pfam: PF01336; tRNA\_antl. 1.  
 DR PRINTS: PRO1042; TRNASYNTHASP.  
 DR PRINTS: PRO0982; TRNASYNTHASP.  
 DR PROSITE: PS00179; AA\_TRNA\_LIGASE\_II\_1; 1.  
 DR PROSITE: PS00339; AA\_TRNA\_LIGASE\_II\_2; UNKNOWN.1.  
 DR ATP-binding; Aminoacyl-tRNA synthetase; Complete proteome; Ligase;  
 KW Protein biosynthesis.  
 SQ SEQUENCE 583 AA; 66166 MW; A3723A2E100222E2 CRC64;

Query Match 44.6%; Score 45; DB 16; Length 583;  
 Best Local Similarity 69.2%; Pred. No. 96;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 8 DAESFGAIAHGGL 20  
 DB 514 DALSGAPPHGSI 526

RESULT 11  
 O9HPT7 PRELIMINARY; PRT; 746 AA.  
 AC O9HPT7.  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE VNG1476C.  
 GN VNG1476C.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 CC Halobacterium  
 NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahatras G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Shroga J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Weller R., Goo Y.A.,  
 RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angvine C.M., Dale H.,  
 RA Isehnager T.A., Peck R.F., Pohlshroder M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 DR EMBL: AE005063; AAC19780.1; -  
 DR InterPro: IPR000731; HMGCR\_patched\_57M.  
 DR PROSITE: PS50156; SSD; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 746 AA; 79248 MW; F278AC41A19B9555 CRC64;

Query Match 44.6%; Score 45; DB 17; Length 746;  
 Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 3 AAVMGDAESFGAIAHGGL 20  
 DB 220 AVEVGVIYTFGAIGYCGI 237

RESULT 12  
 O46426 PRELIMINARY; PRT; 178 AA.  
 AC O46426;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE AOUAPORIN 4 (FRAGMENT).  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NEW ZEALAND;  
 RA Carter E.P., Uemishi F., Matthey M.A., Verkmann A.S.;  
 RT "Increased water permeability across the blood-gas barrier in rabbit  
 lungs in the first 24 hours after birth."  
 RL J. Clin. Invest. 0:0-0(1997).  
 DR EMBL: AF000312; AAB94409.1; -  
 DR HSP; P29972; LFQY.  
 DR InterPro: IPR000425; MIP.  
 DR Pfam: PF00230; MIP; 1.  
 DR PRINTS: PRO0783; MINTRENSICP.  
 KW Porin.  
 FT NON\_TER 1 1  
 FT NON\_TER 178 178  
 SQ SEQUENCE 178 AA; 19236 MW; CE43F9E1BACBAE2 CRC64;

Query Match 44.1%; Score 44.5; DB 6; Length 178;  
 Best Local Similarity 39.3%; Pred. No. 30;  
 Matches 11; Conservative 4; Mismatches 4; Indels 9; Gaps 1;

OY 2 PAVVMGDAES-----FGAIAHGGL 20  
 DB 82 PAVIMGNNENHWIYWGPIGAVLAGGL 109

RESULT 13  
 O92335 PRELIMINARY; PRT; 302 AA.  
 AC O92335;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE AOUAPORIN 4 ISOFORM 1.  
 OS Dipodomys merriami (Merriam's kangaroo rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Heteromyidae;  
 OC Dipodomysinae; Dipodomys.  
 NCBI\_TaxID=94247;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Fang P.K., Huang Y., Walsberg G.E., Brown D., van Hoek A.N.;  
 RT "Molecular Cloning and Characterization of Merriam's Kangaroo Rat  
 Aquaporin 4";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY032857; AAK66823.1; -  
 KW Porin.  
 SQ SEQUENCE 302 AA; 32439 MW; EECDB5A25799CD2 CRC64;

Query Match 44.1%; Score 44.5; DB 11; Length 302;  
 Best Local Similarity 39.3%; Pred. No. 55;  
 Matches 11; Conservative 4; Mismatches 4; Indels 9; Gaps 1;

OY 2 PAVVMGDAES-----FGAIAHGGL 20  
 DB 198 PAVIMGNNENHWIYWGPIGAVLAGGL 225

RESULT 14  
 O92334 PRELIMINARY; PRT; 324 AA.  
 AC O92334;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE AOUAPORIN 4 ISOFORM 2.  
 OS Dipodomys merriami (Merriam's kangaroo rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Heteromyidae;

OC Dipodomysinae; Dipodomys.  
 OX NCBI\_TaxID=94247;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Pang P.K., Huang Y., Malsberg G.E., Brown D., van Hoek A.N.;  
 RT "Molecular Cloning and Characterization of Merriam's Kangaroo Rat  
 Aquaporin 4.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY032858; AAK66824.1; -  
 KW Forin.  
 SQ SEQUENCE 324 AA; 34939 MW; EAE9CE3B17952CDD CRC64;

Query Match 44.1%; Score 44.5; DB 11; Length 324;  
 Best Local Similarity 39.3%; Pred. No. 59;  
 Matches 11; Conservative 4; Mismatches 4; Indels 9; Gaps 1;

OY 2 PAVMGDAES-----FGAIAHGCL 20  
 DB 220 PAVMGWENHWIYWGPIIGAVLAGCL 247

## RESULT 15

O9JZ15 PRELIMINARY; PRT; 469 AA.  
 AC O9JZ15;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE 3-ISOPROPYLMALATE DEHYDRATASE, LARGE SUBUNIT.  
 GN NMB1036.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / SEROGROUP B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
 RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,  
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciltone H., Clark E.B.,  
 RA Cotton M.D., Ulteback T.R., Khouri H., Qin H., Yamathavan J.,  
 RA Gill J., Scariato V., Maignani V., Pizze M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 MC58.";  
 RL Science 287:1809-1815(2000).  
 DR EMBL; AE002454; AAF41435.1; -.  
 DR TIGR; NMB1036; -.  
 DR InterPro; IPR001030; Aconitase.  
 DR Pfam; PF00330; aconitase; 1.  
 DR PRINTS; PR00415; ACONITASE.  
 DR ProDom; PD000511; ACONITASE.  
 DR PROSITE; PS00450; ACONITASE\_1; 1.  
 DR PROSITE; PS01244; ACONITASE\_2; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 469 AA; 50843 MW; D538C334E590E4E0 CRC64;

Query Match 44.1%; Score 44.5; DB 16; Length 469;  
 Best Local Similarity 50.0%; Pred. No. 90;  
 Matches 10; Conservative 3; Mismatches 2; Indels 5; Gaps 1;

OY 4 VVMGDAE-----SGAIAHG 18  
 DB 126 VVCGDSHTSTHGAFCALAHG 145

Search completed: July 8, 2002, 08:21:05  
 Job time: 220 sec

Mon Jul 8 09:33:19 2002

us-09-582-296-2.rspt

